

UNIVERSITÀ DEGLI STUDI DI BARI
FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
CORSO DI LAUREA IN FISICA

TESI DI LAUREA

LA SELEZIONE DELLA
CARATTERISTICA MULTIDIMENSIONALE
NEI SEGNALI FISIOLGICI

RELATORI:

CHIAR.MO PROF. M. PELLICORO

CHIAR.MO PROF. S. STRAMAGLIA

LAUREANDA:

TERESA CREANZA

ANNO ACCADEMICO 2005-2006

alla mia famiglia
e a Lorenzo

Indice

Introduzione	1
1 Le serie temporali fisiologiche	6
1.1 Il controllo a breve termine del sistema cardiovascolare	8
1.2 I dati fisiologici	12
1.3 I segnali di variabilità cardiovascolare	13
1.3.1 La serie RRI	13
1.3.2 Le serie SAP e DAP	14
1.3.3 La serie ILV	16
1.3.4 Pre-trattamento dei segnali	17
2 Le variabili: gli indicatori quantitativi della patologia	20
2.1 I metodi d'analisi delle serie di variabilità cardiovascolare	20
2.1.1 Misure lineari	21

2.1.2	Misure non lineari	22
2.2	Entropia e informazione	25
2.2.1	L'entropia dei processi stocastici	25
2.2.2	Entropia dei sistemi dinamici	28
2.2.3	Entropia delle serie temporali reali	33
2.2.4	L'entropia approssimata e le serie fisiologiche	37
2.2.5	Il metodo MSE	38
2.3	Modello autoregressivo multivariato:VAR	41
2.3.1	Il modello AR	42
2.3.2	Il modello VAR	44
2.4	L'interpretazione dei coefficienti	46
3	La classificazione automatica e la selezione di variabili	50
3.1	La classificazione delle osservazioni	50
3.2	Funzione discriminante lineare di Fisher	52
3.3	Tecniche di selezione	58
3.4	Il criterio di selezione	61
3.5	Analisi discriminante lineare regolarizzata	67

4	L'analisi dei risultati	71
4.1	Le capacità discriminanti individuali delle variabili	72
4.2	Strategie di selezione	74
4.3	Metodo di leave-one-out	81
	Conclusioni	86
A	Le curve R.O.C.	89
B	Il test di Wilcoxon-Mann-Whitney della somma dei ranghi	94
C	Tabelle	97
	Bibliografia	108
	Ringraziamenti	112

Introduzione

Fondamentale nella ricerca medica è lo studio delle serie temporali fisiologiche per comprendere la dinamica dei sistemi di regolazione fisiologica e cercare di sviluppare metodi sempre più raffinati volti all'identificazione di pazienti con varie patologie.

A partire dalle serie temporali si possono ricavare diversi indicatori, o caratteristiche (*feature*), che forniscono informazioni parziali e complementari sulla dinamica di tali sistemi.

Perché queste misure siano utili nella pratica clinica della diagnosi di stati patologici e nella fisiopatologia, è fondamentale applicare tecniche di classificazione automatica, il cui studio costituisce la disciplina detta *pattern recognition*. Classificare significa decidere sulla base degli indicatori calcolati, circa l'appartenenza di un individuo alla classe dei sani o dei malati. Perché il processo sia automatizzabile, è necessario effettuare una serie di misurazioni di tali caratteristiche sui campioni (*feature extraction*), in modo da avere a disposizione dei vettori numerici, visti come realizzazione di un vettore casuale, facilmente manipolabili dal calcolatore, che fornisce in uscita la classe di appartenenza stimata.

In questo lavoro di tesi è stato implementato un classificatore binario basato sull'analisi discriminante e sono state sviluppate diverse procedure di selezione del sottoinsieme delle caratteristiche più significative per stimare l'appartenenza di un individuo alla categoria dei sani o a quella dei malati.

La selezione delle feature più significative è importante per chiarire la fisiopatologia e per rendere più facile l'applicabilità clinica della procedura riducendo con il numero delle variabili anche i costi delle misure. La restrizione del numero di variabili si rivela infine necessaria se è disponibile una casistica limitata per la fase di addestramento. Si dimostra infatti che, a parità di tasso di errore, un classificatore necessita di una sequenza di training che cresce esponenzialmente con il numero delle variabili [1].

In questo lavoro di tesi sono state confrontate le serie temporali *short-term* di variabilità cardiovascolare di soggetti sani con le stesse riferite a individui malati di scompenso cardiaco congestizio CHF (chronic heart failure), un disagio caratterizzato da una degenerazione delle funzioni del sistema autonomico e da anomalie del ventricolo sinistro. I segnali di variabilità cardiovascolare considerati sono le serie di variabilità del ritmo cardiaco RRI (R-R interval), dei massimi e minimi di pressione arteriosa in ogni ciclo cardiaco, SAP (systolic arterial pressure) e DAP (diastolic arterial pressure) rispettivamente, e del volume polmonare istantaneo in corrispondenza dei battiti ILV (instantaneous lung volume) sotto due diversi regimi di respirazione. I metodi e gli strumenti utilizzati per l'acquisizione e il pre-trattamento dei dati sono illustrati nel primo capitolo. Nello stesso capitolo sono inoltre brevemente descritti i meccanismi fisiologici che regolano il controllo a breve termine del sistema cardiovascolare con particolare attenzione al sistema nervoso autonomo SNA e al baroriflesso. Da questo quadro emerge chiaramente che il sistema cardiovascolare è un sistema complesso regolato da meccanismi di interazione che operano attraverso scale temporali e spaziali multiple. Le variabili di uscita di questi sistemi mostrano delle complesse fluttuazioni che non possono essere attribuite semplicemente alla presenza di rumore ma suggeriscono la caoticità della dinamica sottostante.

Sono stati fatti molti studi per determinare la natura caotica dell'attività cardiaca applicando metodi di analisi della teoria dei sistemi non lineari [2]. Tuttavia sono

emerse delle difficoltà fondamentali nell'implementazione di questi metodi data la natura rumorosa dei segnali biologici, la lunghezza finita delle serie di dati e la loro non stazionarietà.

Dati questi problemi, gli sforzi dei ricercatori si sono orientati allo sviluppo di misure e tecniche di analisi dei dati che non richiedono alcuna ipotesi sulla natura del sistema (dinamico deterministico caotico o stocastico). Tali sono alcuni algoritmi che cercano di distinguere i segnali in base alla loro complessità, irregolarità e casualità piuttosto che alla loro natura caotica. Una delle feature usate per il nostro classificatore è appunto una misura di complessità.

Diciamo subito che non esiste una definizione formale e unica della complessità [3]: ne esistono diverse definizioni operative ognuna con un limitato dominio di applicabilità. In generale, la complessità è correlata alla comprensione, cioè alla capacità di un soggetto di fornire una descrizione chiara ma breve di un sistema. Essa è quindi una funzione del soggetto e dell'oggetto: il sistema, o una sua rappresentazione adatta, deve essere convenientemente diviso in parti a loro volta divise in sottoparti, producendo così una gerarchia. A questo punto il soggetto deve studiare le interazioni tra i sottosistemi e incorporarli nel modello. Considerare le interazioni a diversi livelli di risoluzione conduce al concetto di *scaling* [4].

L'algoritmo da noi considerato è il metodo dell'entropia multiscala MSE, una misura di predicibilità delle serie temporali considerate a diverse scale temporali ottenute una dall'altra mediante una procedura di *coarse graining*.

Nel secondo capitolo, pertanto, è esposta una rassegna di diverse misure di entropia che parte dalle definizioni storiche dell'entropia di Shannon e Kolmogorov-Sinai che ne costituiscono le basi e include l'entropia approssimata introdotta da Pincus, più adatta all'applicazione ai sistemi fisiologici e in particolare alle serie *short-term*. Infine è stata introdotta la statistica inglobata nell'algoritmo di MSE, ovvero l'entropia

di campionamento SE (*sample entropy*).

Dopo aver quantificato l'imprevedibilità delle serie alle diverse scale, si può pensare ad un modello del sistema che tenga conto delle sorgenti di imprevedibilità e dello scaling temporale. Abbiamo già detto che la complessità delle fluttuazioni suggerisce tra le sorgenti di imprevedibilità la dinamica caotica oltre ad una componente di rumore esterno, ma un modello che tenga conto in maniera esatta delle due componenti non è applicabile [5]. E' possibile assumere che la nonlinearietà sia una piccola perturbazione di un sistema essenzialmente stocastico lineare, o, alternativamente, che l'elemento di rumore esterno sia una piccola perturbazione di un processo deterministico non lineare. Seguendo il primo approccio e tenendo conto dell'accoppiamento tra le diverse serie di variabilità cardiovascolare tramite i diversi meccanismi di regolazione è stato costruito un modello autoregressivo vettoriale VAR alle diverse scale temporali. Nel secondo capitolo è spiegato come sono stati calcolati i coefficienti del VAR che costituiscono, insieme alle entropie SE alle diverse scale, le feature del nostro classificatore.

Nel terzo capitolo, dopo un'introduzione formale alla teoria della classificazione, è stato esaminato un metodo di riconoscimento automatico basato su tecniche di statistica multivariata: l'analisi discriminante di Fisher. In particolare, abbiamo dimostrato che nell'ipotesi di distribuzioni normali con uguali matrici di dispersione si può determinare la funzione discriminante lineare di Fisher con il metodo dei minimi quadrati ordinari o, nel caso di problemi mal posti nel senso di Hadamard, regolarizzati. Nello stesso capitolo abbiamo esposto le diverse procedure di selezione automatica delle feature più significative nella classificazione: la selezione in avanti (forward selection), l'eliminazione all'indietro (backward elimination) e l'algoritmo Recursive Feature Elimination (RFE) di Guyon [6].

Nel quarto capitolo abbiamo discusso i valori delle entropie e dei coefficienti del VAR

calcolati per le nostre serie e confrontato la loro individuale capacità discriminante tra sani e malati mediante test statistici univariati (test di Wilcoxon-Mann-Whitney della somma dei ranghi, area ROC, distanza di Mahalanobis). Nello stesso capitolo abbiamo implementato il nostro classificatore che usa i metodi dei minimi quadrati regolarizzati per trovare la funzione discriminante di Fisher scegliendo come parametro di regolarizzazione quello che minimizza l'errore di leave-one-out di predizione. Abbiamo applicato le diverse procedure di selezione suddette per individuare per ogni dimensione possibile il classificatore migliore. In base ad ogni strategia abbiamo scelto la cardinalità e gli elementi del sottoinsieme di feature migliore e infine abbiamo scelto, tra questi, il sottoinsieme ottimo in base al criterio di massima semplicità. Il sottoinsieme con 96 feature migliore secondo la procedura di RFE risulta il sottoinsieme di dimensione minima compreso nella barra d'errore standard del sottoinsieme con in assoluto il valore più piccolo d'errore, quindi il nostro classificatore globalmente ottimo.

Ringraziamo il Dipartimento di Bioingegneria e Tecnologie Biomediche, Fondazione Salvatore Maugeri, IRCSS Istituto Scientifico di Montescano (PV) nelle persone di R. Maestri e G. D. Pinna per aver messo a disposizione i dati fisiologici analizzati.

Capitolo 1

Le serie temporali fisiologiche

L'analisi della variabilità del periodo cardiaco e della pressione arteriosa, cioè la variabilità cardiovascolare, è stata usata nelle ultime due decadi come strumento quantitativo non invasivo per valutare il funzionamento del sistema nervoso autonomo SNA.

Uno studio di Akselrod et al. apparso sulla rivista Science nel 1981 dimostrava per la prima volta in maniera chiara che lo studio delle serie di variabilità del ritmo cardiaco RRI può fornire in modo non invasivo informazioni sul controllo simpato-vagale del cuore [7]. Oggi alcuni aspetti della relazione tra la variazione del ritmo cardiaco e il controllo neurale del sistema cardiovascolare sono chiari: la varianza ridotta della serie RRI è un fattore prognostico sfavorevole nei pazienti con scompenso cardiaco congestizio CHF (chronic heart failure) ed è da attribuire ad una riduzione della modulazione vagale, simpatica e delle loro interazioni. In particolare indici di sbilanciamento autonomo e di attivazione neuro-ormonale forniscono informazione prognostica sulla frazione di emissione ventricolare sinistra e predicono il rischio di mortalità precoce. Conoscere meglio il meccanismo che genera la disfunzione autono-

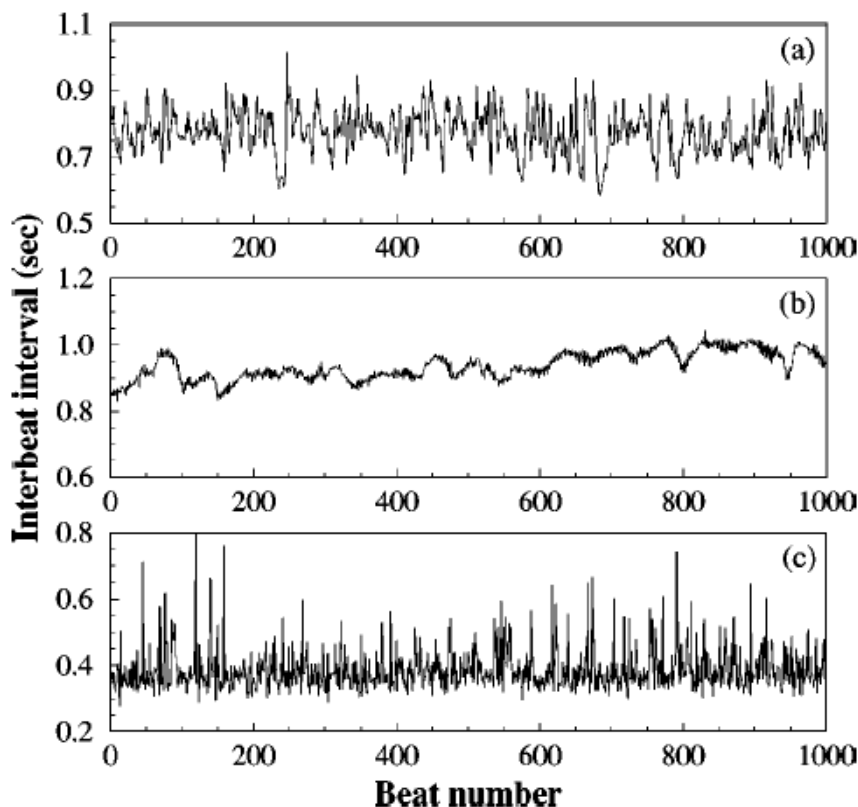


Figura 0.1: Serie temporali RRI rappresentative relative ad a) un individuo sano, b) un soggetto con CHF, c) un soggetto con fibrillazione atriale, un' aritmia cardiaca altamente erratica [8]

ma in soggetti con CHF appare ancora più essenziale se si pensa che tale disfunzione gioca un ruolo chiave nella progressione della malattia.

Anche la modulazione del ritmo cardiaco ad opera dei baroriflessi è ridotta nell'insufficienza cardiaca. Questo deriva dalla combinazione di una serie di fattori come un'anormalità a livello cerebrale, uno shift dell'interazione simpato-vagale verso l'eccitazione simpatica e l'alterazione di parecchi siti di input nel sistema periferico ([9]). Nel prossimo paragrafo sono descritti brevemente i meccanismi di regolazione del sistema cardiovascolare. Da questa descrizione emergono chiaramente la correlazione

tra le variazioni del ritmo cardiaco e della pressione arteriosa attraverso il sistema nervoso autonomo SNA e il baroriflesso e l'influenza su di loro dell'attività respiratoria direttamente (alla frequenza respiratoria, intorno ai 0.25 Hz) e attraverso il loro ciclo di controllo (intorno a 0.1 Hz). Questo giustifica la scelta di considerare le serie RRI, SAP, DAP e ILV per monitorare il funzionamento del sistema cardiovascolare in soggetti CHF mediante il confronto con quelle di individui sani.

1.1 Il controllo a breve termine del sistema cardiovascolare

Lo scopo del sistema cardiovascolare è fornire il sangue ai tessuti e agli organi vitali sotto tutte le condizioni. Al pari di altri importanti sistemi fisiologici, esso è caratterizzato da complessi meccanismi di controllo volti al mantenimento dell'omeostasi, cioè a garantire che alcune variabili vitali (concentrazione di ossigeno nel sangue, temperatura corporea, ...) rimangano entro stretti limiti di tolleranza.

Nel caso del controllo del ritmo cardiaco e della pressione arteriosa, questi meccanismi devono essere in grado, da un lato di reagire prontamente a stimoli di diversa natura ripristinando l'equilibrio per mezzo di regolazioni a breve termine, dall'altro di riadattarsi in risposta a variazioni lente quali quelle indotte dall'invecchiamento o da particolari patologie. Noi consideriamo solo il controllo a breve termine del sistema cardiovascolare: cioè il controllo di meccanismi che agiscono con un ritardo che va da pochi millisecondi ad alcuni minuti.

Il cuore è l'organo centrale del sistema cardiovascolare che ha il compito di pompare il sangue ricco di ossigeno e altre sostanze nutritive alla cute e agli organi vitali (tra cui reni, cervello, fegato) attraverso una fitta rete di arterie e capillari. Sebbene la

ritmicità intrinseca del cuore è determinata dal nodo seno-atriale SA, il *pacemaker* naturale localizzato nell'atrio destro, il ritmo è continuamente modulato dal sistema nervoso autonomo SNA mediante impulsi nervosi simpatici e parasimpatici mandati dal cervello al nodo SA. Il ritmo finale del cuore è stabilito dal bilanciamento tra gli impulsi nervosi simpatici che tendono ad accelerarlo e quelli vagali che tendono a rallentarlo. I tempi di latenza delle branche simpatiche e parasimpatiche differiscono significativamente: le prime hanno tipicamente un tempo di risposta di pochi secondi, le altre hanno una risposta agli stimoli quasi immediata.

La pressione arteriosa è una conseguenza della contrazione del cuore che forza il volume sistolico SV (stroke volume) ¹nell'aorta. A causa della resistenza vascolare questo flusso iniziale causa la pressione arteriosa all'interno dei vasi sanguigni. Quest'azione di pompaggio del cuore insieme alla resistenza periferica totale TPR mantiene una continua differenza di pressione tra il sito arterioso e quello venoso del sistema vascolare, prerequisito essenziale per l'irrogazione dei tessuti. La pressione arteriosa è il risultato del prodotto di resistenza periferica, frequenza cardiaca e volume sistolico (effetto Windkessel). Ciò che principalmente determina il SV è il grado di riempimento del cuore e ,quindi, il ritorno venoso: meccanismo di Frank-Starling. Quando il riempimento diastolico del cuore aumenta di un dato volume, il volume sistolico aumenta della stessa quantità. La TPR è controllata mediante la vasocostrizione e la vasodilatazione principalmente dalle branche simpatiche. Attraverso il controllo locale della resistenza vascolare si ottiene la regolazione della distribuzione del flusso sanguigno tra i diversi tessuti. Il sistema SNA è un sistema di controllo a feedback. Sebbene l'integrazione centrale controlla il comportamento globale, parecchi riflessi forniscono rapidi meccanismi di feedback per rispondere efficacemente alle specifiche richieste. Il più importante di questi riflessi è il baroriflesso: esso è basato sui

¹ Il volume sistolico o gittata sistolica è il volume di sangue espulso ad ogni sistole

barocettori (recettori di pressione) che attraverso la modulazione vagale e simpatica inducono variazioni di frequenza cardiaca, contrattilità e resistenze periferiche in risposta a rapide variazioni pressorie. Le terminazioni nervose sensitive del riflesso barocettivo sono localizzate nella parete del seno carotideo e dell'arco aortico e nella parete di altri grossi vasi sanguigni. I barocettori arteriosi, stimolati da un aumento della pressione sono in grado di inibire l'attività efferente simpatica e allo stesso tempo di aumentare l'attività vagale con conseguente riduzione della frequenza cardiaca, della contrattilità e della vasocostrizione sistemica. Al contrario una riduzione della pressione arteriosa provoca una diminuzione della frequenza di scarica dei barocettori con conseguente attivazione simpatica e inibizione vagale. Negli esseri umani la latenza del meccanismo baroriflesso è stata valutata tra 200 e 600 ms; quindi questo è un meccanismo di regolazione veloce nello stabilizzare la pressione arteriosa.

Il sistema complessivo opera in circuito chiuso. La pressione sanguigna, risultato del prodotto di resistenza periferica e portata cardiaca ², viene misurata dai barocettori che inviano il segnale al sistema nervoso centrale il quale, attraverso l'azione dei riflessi simpatico e vagale, regola la frequenza imposta dal *pacemaker*. Quest'ultima influenza il riempimento del muscolo cardiaco e quindi la portata (attraverso la legge di Stirling).

Tra gli altri riflessi cardiovascolari abbiamo i barorecettori cardiopolmonari e i chemocettori sensibili alle basse pressioni; inoltre, il controllo renale del volume del sangue, il sistema renina-angiotensina-aldosterone e la termoregolazione possono giocare un ruolo nella regolazione del sistema cardiovascolare a breve termine.

In soggetti normali sani la frequenza cardiaca aumenta (decrece) e la pressione arteriosa decresce (aumenta) durante la fase inspiratoria (nella fase espiratoria). L'attività respiratoria influenza la variabilità cardiovascolare in diversi modi. Il ritmo respira-

² la portata cardiaca è la quantità di sangue pompata in media dal cuore in un minuto (circa 5 l/min); corrisponde quindi al volume sistolico per la frequenza

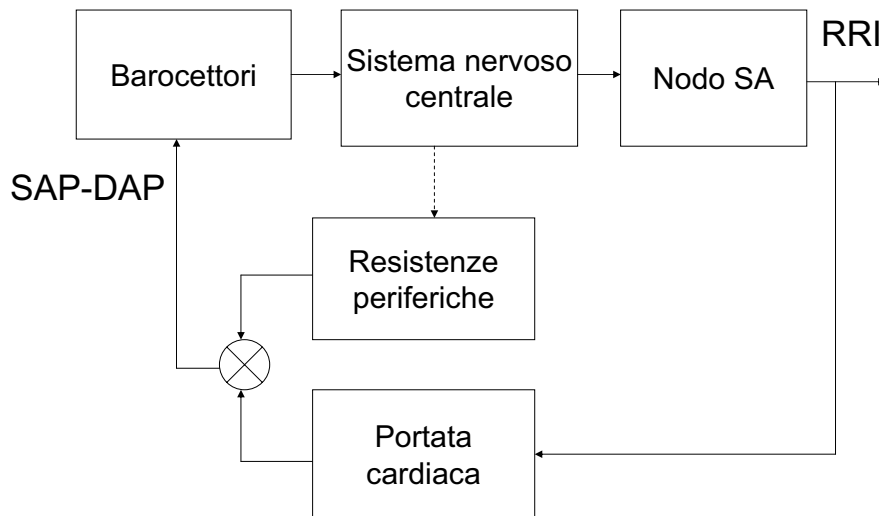


Figura 1.2: Schema del loop chiuso di regolazione della pressione e del ritmo cardiaco torio, generato nel tronco dell'encefalo, modula sia l'attività efferente simpatica e parasimpatica verso il cuore e i vasi attraverso l'accoppiamento diretto tra centri respiratori e centri autonomi che la sensibilità dei barorecettori e di altri input afferenti. La modulazione efferente del ritmo cardiaco mediante la respirazione è mediata sia dalle branche simpatiche che da quelle parasimpatiche del SNA, le prime essendo dominanti alle frequenze respiratorie maggiori di 0.15 Hz mentre l'importanza relativa delle altre cresce col decrescere della frequenza respiratoria. A livello meccanico diretto i cambiamenti di pressione intratoracica modulano le pressioni e i flussi arteriosi e venosi toracici e influenzano il ritorno venoso che a sua volta influenza la gittata cardiaca e quindi la pressione arteriosa. Inoltre, lo stiramento meccanico diretto del nodo SA e dei barocettori durante la respirazione possono giocare un ruolo nel modulare RRI. L'accoppiamento ciclo chiuso tra RRI e la pressione arteriosa PA può

anche influenzare la loro variabilità respiratoria. La modulazione respiratoria del ritmo cardiaco (respiratory sinus arrhythmia RSA) agisce sulla PA dato che alterazioni dell'intervallo RR influenzano i valori di PA risultanti dal meccanismo di Frank-Starling (feedward path), mentre il baroriflesso cambia a sua volta RR in seguito al cambiamento di PA.

1.2 I dati fisiologici

I dati analizzati sono stati messi a disposizione dall'Unità Insufficienze Cardiache dell'Istituto Scientifico di Montescano, Fondazione Salvatore Maugeri (Pavia). Abbiamo analizzato dati relativi a 47 volontari sani di età 53 ± 8 anni (media \pm deviazione standard), e 256 pazienti di età 52 ± 9 anni (media \pm deviazione standard) affetti da scompenso cardiaco congestizio CHF (congestive heart failure), una malattia per cui è alta la possibilità di morte cardiaca improvvisa. Tale scompenso è dovuto principalmente a miocardiopatie dilatative (malattia del tessuto muscolare cardiaco) di natura ischemica (ossia causate da un'ostruzione dell'afflusso di sangue o da vasocostrizione) per il 48% dei pazienti esaminati e di natura idiopatica (vale a dire di natura spontanea o di causa ignota) per il 44% degli esaminati. Durante i tre anni successivi alla presa dei dati 47 pazienti sono deceduti per problemi cardiaci. Il protocollo sperimentale prevedeva la registrazione di una traccia elettrocardiografica ECG, del volume polmonare istantaneo ILV con il RIP (Respirace Plus) e della pressione arteriosa PA con il dispositivo Finapres 2300:

- per circa 8 minuti in posizione supina e in regime di respiro basale, (il soggetto respira in maniera spontanea)
- per altri 8 minuti in posizione supina e in regime di respiro controllato, (il

soggetto segue una voce umana registrata che induce fasi di inspirazione ed espirazione alla frequenza di 0.25 Hz).

La respirazione controllata è una pratica usata per regolarizzare e standardizzare l'attività respiratoria durante le ricerche sulle funzioni di controllo autonomo sul sistema cardiovascolare.

Nel paragrafo seguente spiegheremo meglio cosa rappresentano queste misure, descriveremo gli strumenti di registrazione e il pretrattamento dei dati.

1.3 I segnali di variabilità cardiovascolare

1.3.1 La serie RRI

L'elettrocardiogramma registra l'attività elettrochimica del cuore che è alla base del meccanismo di pompaggio del sangue. Il muscolo cardiaco può essere considerato come un mezzo eccitabile spazialmente esteso con una fase eccitabile, una eccitata ed una refrattaria. All'inizio di un ciclo cardiaco, un segnale è generato al nodo seno-atriale, un insieme di cellule muscolari. L'onda di depolarizzazione risultante procede lungo un ben definito cammino, prima attraverso gli atri (onda P) e poi attraverso i ventricoli (complesso QRS). Il tessuto eccitato si contrae e quindi emette sangue verso il corpo e i polmoni. Infine, tutto il tessuto cardiaco risulta eccitato ed è refrattario finché lo stimolo si esaurisce. Si ha la ripolarizzazione ventricolare che corrisponde al rilassamento dei ventricoli (onda T).

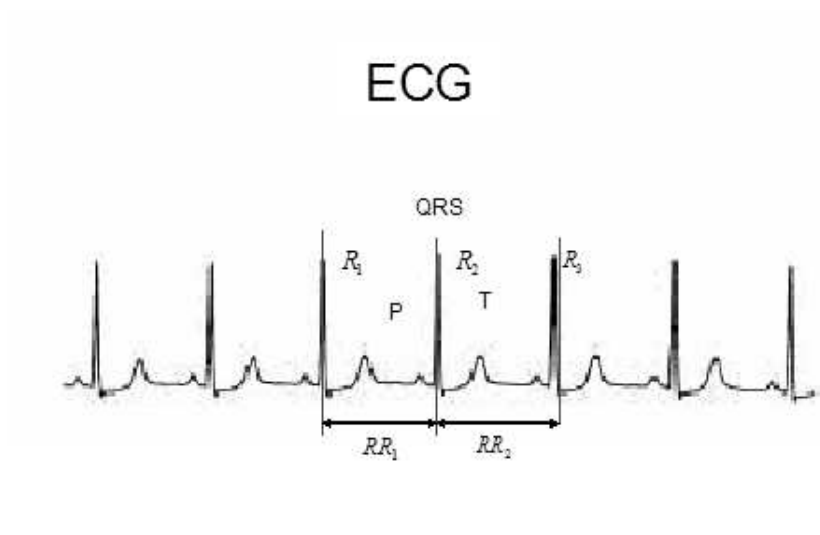
Il cammino dell'onda di depolarizzazione non varia molto da ciclo a ciclo, tuttavia l'istante in cui inizia un nuovo ciclo fluttua in modo abbastanza incomprensibile. L'intervallo di tempo che intercorre tra due picchi R successivi (tra due successive

depolarizzazioni ventricolari) viene indicato con l'acronimo RRI (intervallo R-R) ed equivale alla durata di un intero ciclo cardiaco ventricolare, ossia il tempo che trascorre tra due battiti successivi. Partendo dall'ECG vengono riconosciuti i picchi R e riportati i valori degli intervalli RR in funzione del numero d'ordine del battito: questa serie costituisce il tacogramma.

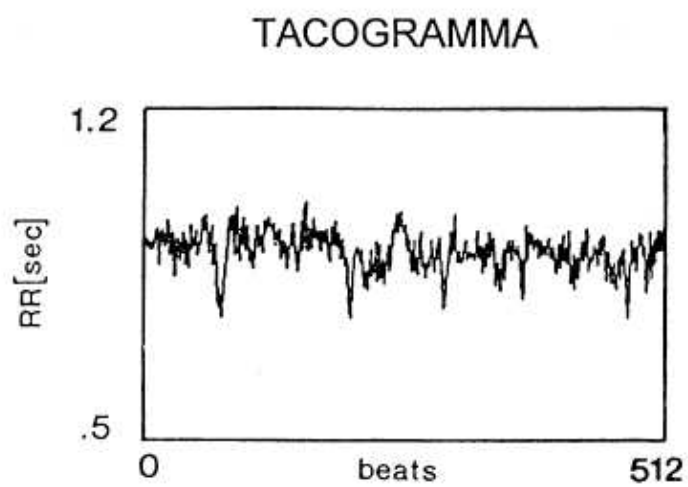
1.3.2 Le serie SAP e DAP

La pressione arteriosa è la pressione che il sangue esercita sulle pareti delle arterie che varia in relazione alle fasi dell'attività cardiaca: è maggiore quando il cuore si contrae (fase di sistole), minore quando esso si rilassa (fase di diastole). Precisamente, durante ogni ciclo, in corrispondenza dell'onda R la pressione raggiunge un valore massimo (pressione arteriosa sistolica SAP), poi la pressione decresce fino ad un valore minimo (pressione arteriosa diastolica DAP), raggiunto immediatamente prima del battito del cuore. Si può registrare un segnale di pressione arteriosa PA continuo in maniera non-invasiva con dispositivi Finapres, di solito il segnale è poi parametrizzato in serie temporali battito a battito dei valori di SAP e di DAP durante ogni ciclo.

I nostri valori di PA sono stati registrati con il dispositivo "Pressione Sanguigna Automatica" (ABP) Finapres il cui funzionamento si basa sulla tecnica di Penaz (1973) definita "volume clamp" (volume fissato). Il principio di funzionamento è quello di mantenere costante il diametro di un'arteria nonostante i cambiamenti di pressione durante il battito cardiaco; questo viene attuato con una cuffia pneumatica, dotata di pletismografo, avvolta intorno al dito. Il pletismografo è composto da una sorgente LED infrarossa e da un detettore: il segnale prodotto è proporzionale all'assorbimento dei raggi infrarossi da parte di ciò che si trova tra sorgente e detettore: durante la fase di sistole, la maggior quantità di sangue che affluisce al dito e che crea un



(a) elettrocardiogramma



(b) tacogramma relativo

Figura 3.3: Elettrocardiogramma rappresentativo e relativo tacogramma

aumento del diametro dell'arteria, comporta un maggior assorbimento e quindi una minore rilevazione dell'infrarosso da parte del detettore, nella fase di diastole si ha, al contrario, un aumento del segnale rilevato. Quando viene rilevato dal pletismografo un aumento di diametro dell'arteria, un sistema automatico aumenta immediatamente la pressione nella cuffia per contrastare questo incremento. Viceversa, quando il diametro si riduce in fase di diastole, la cuffia viene automaticamente sgonfiata. In questo modo l'arteria è bloccata ad un certo diametro definito "set point". Se tale diametro corrisponde allo stato scarico della parete, cioè quello per cui la pressione trasmurale è zero, la pressione della cuffia è in ogni istante pari a quella interna. Il dispositivo è dotato di correzione dell'altezza idrostatica per compensare la componente idrostatica della pressione sanguigna nel dito quando questo non si trova all'altezza del cuore. I segnali di volume e pressione ottenuti sono poi convogliati ad un microprocessore per controllare i battiti e acquisire i dati.

1.3.3 La serie ILV

L'attività respiratoria è stata monitorata mediante la misura del volume istantaneo polmonare ILV (instantaneous lung volume) effettuata con il Respirace Plus, un pletismografo respiratorio ad induttanza. Esso consiste di due bande di tessuto elastico avvolte intorno al paziente una all'altezza del torace e l'altra all'altezza dell'addome, da un oscillatore e da un demodulatore. Su ogni fascia è cucito a zig-zag un filo conduttore isolato che costituisce una spira elettrica con induttanza proporzionale all'area della sezione trasversale racchiusa. L'oscillatore invia un segnale sinusoidale alla spira e la frequenza del segnale in uscita dall'oscillatore varia al variare dell'induttanza riflettendo i cambiamenti della relativa sezione trasversale in seguito ai movimenti respiratori. Attraverso la misura della variazione delle due sezioni, lo stru-

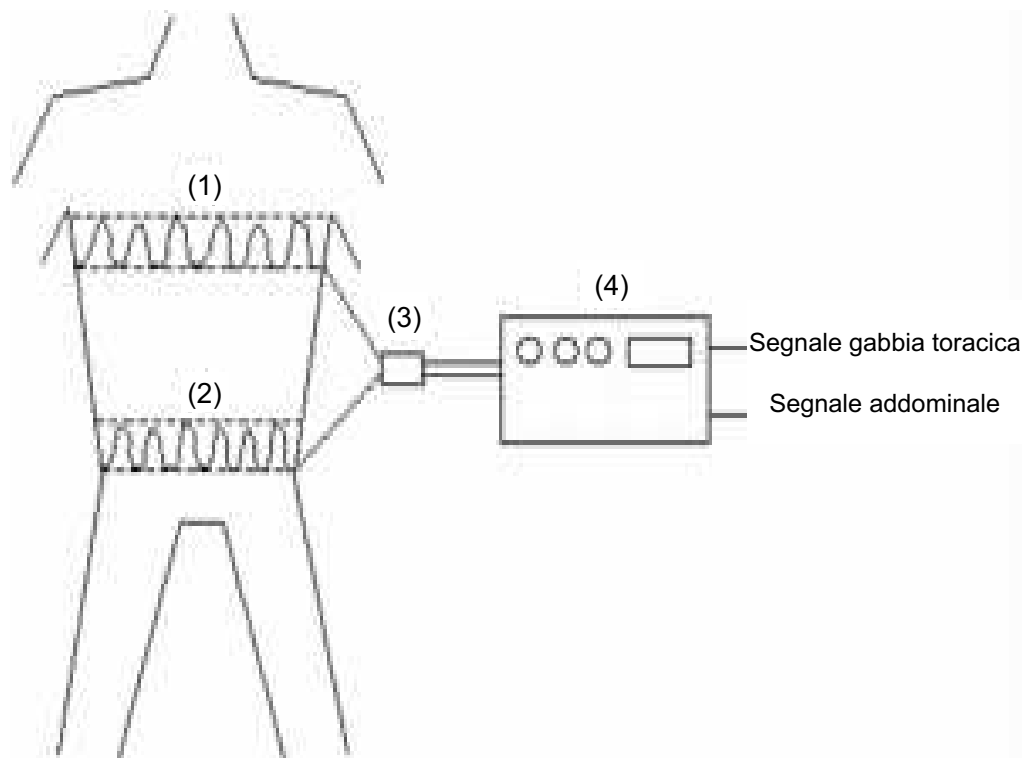


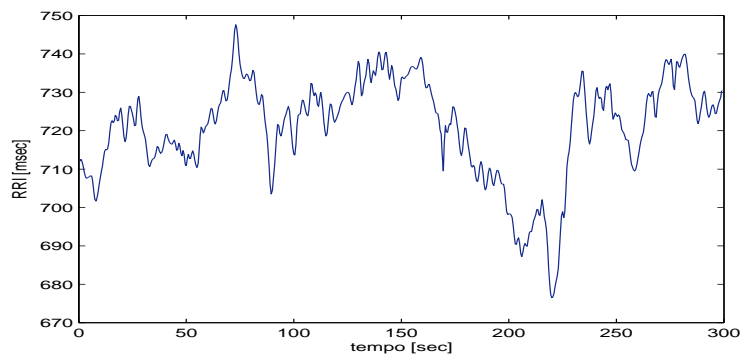
Figura 3.4: Rappresentazioni schematica del set-up sperimentale dove (1) e (2) indicano le fasce elastiche, (3) l'oscillatore, (4) il demodulatore

mento produce una stima del volume polmonare. La serie ILV è la serie dei valori istantanei del volume polmonare in corrispondenza dei picchi R.

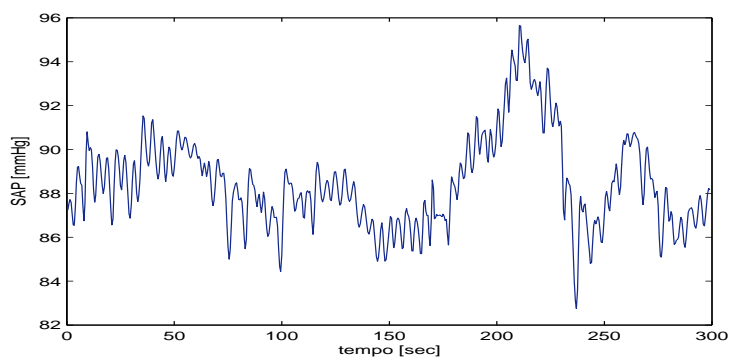
1.3.4 Pre-trattamento dei segnali

Le quattro serie a disposizione sono di differenti tipi: le serie RRI, SAP, DAP sono serie riferite al numero d'ordine del battito mentre la serie ILV è una serie temporale con intervalli di tempo non costanti. Tuttavia per un'analisi congiunta dei quattro segnali abbiamo bisogno di una loro presentazione sincrona.

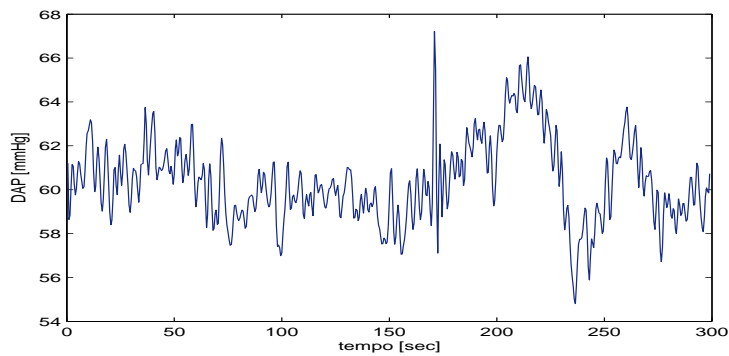
Il valore dell'intervallo RR_n , intervallo tra il picco (n+1)-esimo R_{n+1} e quello n-esimo



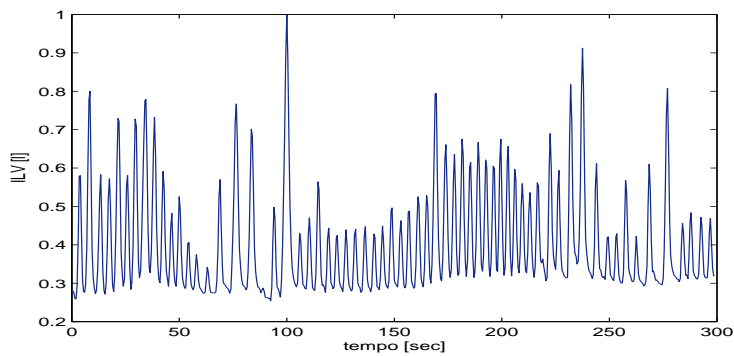
(a) serie RRI



(b) serie SAP



(c) serie DAP



(d) serie ILV

Figura 3.5: Un esempio rappresentativo delle serie analizzate

R_n , viene riscritto in corrispondenza dell'istante in cui si verifica il picco R_{n+1} , mentre in corrispondenza degli istanti nell'intervallo RR_n in cui abbiamo registrato il valore di ECG riportiamo i valori ottenuti per interpolazione cubica tra i valori RR_n e RR_{n-1} . Otteniamo così la serie RRI temporale cubica a tratti con lo stesso tasso di campionamento della serie ECG. Analogamente riscrivendo i valori di SAP, DAP e ILV in corrispondenza dei picchi ed interpolando con un polinomio di terzo grado tra questi punti, otteniamo delle serie campionate con un tasso pari a quello di registrazione dei rispettivi strumenti di acquisizione dati.

A questo punto le quattro serie temporali ottenute vengono ricampionate tutte alla stessa frequenza di 2Hz. E' stata selezionata una sottoregistrazione di 5 minuti con tutti i segnali liberi da artefatti, grandi transienti o grossi cambiamenti nel comportamento delle fluttuazioni.

Capitolo 2

Le variabili: gli indicatori quantitativi della patologia

2.1 I metodi d'analisi delle serie di variabilità cardiovascolare

Una serie temporale non è un modo molto compatto per rappresentare un fenomeno che sta evolvendo nel tempo, è di solito necessario condensare l'informazione in esse contenuta e trovare una parametrizzazione che contenga le caratteristiche più rilevanti per descrivere la dinamica sottostante. La maggior parte dei modi di descrivere quantitativamente una serie temporale derivano da metodi per descrivere la dinamica sottostante ipotizzata.

Il principio consiste nell'assumere che una certa classe di processi ha generato la serie temporale e che da questa si ricava una stima della misura che quantifica il processo [10]. Per esempio, ipotizzando un processo stocastico lineare, per calcolare la media

del processo non conoscendone la distribuzione di probabilità se ne stima il valore a partire dalla serie temporale. I metodi lineari interpretano tutta la struttura regolare in una serie di dati, come le frequenze dominanti o le correlazioni lineari. Dalla teoria del caos deterministico derivano invece una serie di misure adatte all'analisi di serie temporali irregolari.

Kantz, nel suo libro sull'analisi non lineare delle serie temporali [5], introduce diverse di queste misure non lineari esponendone i vantaggi e gli svantaggi rispetto alle più tradizionali misure lineari. Per quanto concerne gli svantaggi è bene sottolineare che le stime di quantità non lineari, a differenza di quelle lineari, di solito non possono essere studiate sufficientemente per provare rigorosamente, ad esempio, la loro correttezza e convergenza. Tra l'altro, come vedremo nei paragrafi seguenti, le informazioni che si possono ottenere da misure non lineari e misure lineari sono complementari ed è quindi consigliabile affiancare i diversi tipi di misure.

Nel seguito è esposta una rassegna delle misure più usate per le serie cardiovascolari e dei risultati più notevoli nello studio dei meccanismi di controllo cardiovascolare nelle più diffuse condizioni fisiopatologiche.

2.1.1 Misure lineari

Le misure lineari più semplici sono la media e la varianza del segnale ma queste, dato che non dipendono dall'ordinamento dei dati, non danno informazione sull'evoluzione temporale del sistema. Tali informazioni si possono ottenere ad esempio dalle autocorrelazioni a diversi ritardi, o equivalentemente, nel dominio delle frequenze, mediante l'analisi dello spettro delle potenze. Diversi studi hanno dimostrato che un valore ridotto di SDNN, una misura della varianza durante 24 ore, è un fattore prognostico negativo indipendente nel periodo dopo un infarto miocardico o nello

scompenso cardiaco congestizio e si ritiene che questa riduzione indichi una riduzione della modulazione vagale simpatica e delle loro interazioni.

Per stimare inoltre la densità spettrale di potenza dei segnali cardiovascolari sono usati di solito due metodi diversi: la trasformata rapida di Fourier o la stima spettrale di un modello autoregressivo la quale fornisce una migliore risoluzione spettrale per serie di lunghezza ridotta. In soggetti sani, in brevi spezzoni di RR, si possono individuare due componenti spettrali principali: quella a bassa frequenza (tra 0.03 e 0.15 Hz) e quella ad alta frequenza (tra 0.15 e 0.4 Hz). E' stato dimostrato che un aumento dell'attività simpatica diretta al cuore, ottenuto con manovre di laboratorio standard, induce un incremento della potenza normalizzata della componente a bassa frequenza. Si è osservata una variazione in senso opposto nella fase avanzata dello scompenso cardiocircolatorio. Non è ancora chiara l'interpretazione fisiopatologica da attribuire alle modifiche delle componenti oscillatorie che si osservano in varie condizioni patologiche [7].

2.1.2 Misure non lineari

Le serie di variabilità cardiovascolare mostrano delle complesse fluttuazioni che non possono essere attribuite semplicemente alla presenza di rumore ma suggeriscono la caoticità della dinamica sottostante. Quando si sospetta che il segnale sia deterministico caotico si deve tenere presente che la quantità d'informazione che si può quantificare con misure lineari è ridotta. Ad esempio tramite l'analisi spettrale i picchi indicano senza dubbio segnali periodici o quasi periodici ma non siamo capaci di distinguere nella parte continua dello spettro il contributo del rumore stocastico da quello del caos deterministico.

Sono stati fatti invece molti studi per determinare la natura caotica dell'attività

cardiaca applicando metodi di analisi della teoria dei sistemi non lineari. Per esempio, ipotizzando la presenza di un attrattore caotico sottostante, sono state stimate le proprietà di scaling della dinamica generante sulla base dell'analisi dimensionale. Nella teoria della dinamica non lineare lo spazio degli stati è ricostruito dalle serie RRI (*embedding*) e il comportamento dinamico della dinamica ricostruita può essere quantificato mediante il massimo esponente di Lyapunov o la dimensione di correlazione. Tuttavia sono emerse delle difficoltà fondamentali nell'implementazione di questi metodi data la natura rumorosa dei segnali fisiologici, la lunghezza finita delle serie di dati e la loro non stazionarietà.

Dati questi problemi, gli sforzi dei ricercatori si sono orientati allo sviluppo di misure non lineari di analisi dei dati che non richiedono alcuna ipotesi sulla natura del sistema (dinamico deterministico caotico o stocastico). Tali sono alcuni algoritmi che cercano di distinguere i segnali in base alla loro complessità, irregolarità e casualità piuttosto che alla loro natura caotica. Essi rappresentano un'applicazione della teoria dell'informazione alla dinamica alto-dimensionale degli organismi viventi. Esempi di questo tipo sono l'entropia approssimata (Pincus), l'entropia rinormalizzata (Kurths et al. 1995), le entropie binarie (Cysarz et al. 2000) e la strategia indicata da Porta et al. (2000) definita "dominio dell'informazione" [2]. Nel paragrafo seguente sarà esposto nel dettaglio l'entropia approssimata di Pincus che è alla base dell'algoritmo MSE utilizzato nell'analisi delle nostre serie.

Sono stati fatti degli studi per evidenziare il legame tra la complessità dell'RRI e il controllo nervoso del cuore mediante l'analisi della dimensione di correlazione, l'entropia di Kolmogorov e gli esponenti di Lyapunov nell'RRI. Nei pazienti dopo recente trapianto cardiaco queste misure risultavano tutte ridotte rispetto ai sani, tuttavia erano ancora presenti esponenti di Lyapunov positivi. Ricordiamo che gli esponenti di Lyapunov misurano il tasso di divergenza di traiettorie inizialmente vicine nello

spazio delle fasi e la positività di questi parametri è indice della presenza di caos deterministico nel segnale. Ho et al. sono stati tra i primi a sostenere che nei pazienti con scompenso cardiaco gli indici di non linearità costituiscono un valido contributo alla capacità prognostica dell'RRI [7].

Accanto alle misure non lineari di complessità esistono misure delle proprietà frattali che riflettono l'invarianza a diverse scale temporali: molti segnali fisiologici non hanno una scala temporale caratteristica e può accadere che il segnale $y(t)$ e la sua versione scalata $y(at)$ abbiano la stessa statistica (media, varianza, etc...).

Una delle misure frattali più usate nell'analisi di serie RRI, [Peng et al., 2000; Mäkikallio, 1998], è il metodo DFA (*detrended fluctuation analysis*). Esso costruisce le serie temporali integrate su blocchi di dimensione n con n via via maggiore. La fluttuazione su ogni blocco è calcolata come la radice quadratica media della differenza tra la serie integrata e il suo miglior fit lineare nel blocco. Tali fluttuazioni seguono una legge di potenza e l'esponente è detto parametro di autosimilarità [11]. Da uno studio di Peng et al.(2000) risulta un valore del parametro di autosimilarità maggiore in pazienti con scompenso cardiaco rispetto allo stesso valore misurato nel gruppo dei sani.

Nel paragrafo seguente è descritta in dettaglio la base teorica e l'implementazione dell'algoritmo MSE (multiscale sample entropy), ideato da Costa, Goldberger e Peng, utilizzato per stimare il grado di complessità e di predicibilità delle nostre serie RRI, SAP, DAP e ILV di breve durata. Notiamo che tale metodo è stato già applicato dagli autori a serie RRI derivate da tracce elettrocardiografiche di 24 ore sotto condizioni fisiologiche e patologiche [8]. Il loro studio conferma i risultati trovati per altra via che mettono in luce una generale perdita di complessità con l'età e in casi di fibrillazione atriale e di CHF.

2.2 Entropia e informazione

La teoria dell'informazione fornisce un'importante strumento per l'analisi delle serie temporali: l'osservazione del sistema è considerata come una sorgente di informazione.

La serie è una stringa di numeri che rappresenta un messaggio trasmesso dal sistema. Conoscendo le distribuzioni di probabilità di questi numeri e le probabilità di transizioni da un numero all'altro, possiamo mediante l'entropia valutare la quantità di informazione immagazzinata nella serie [5]. L'entropia è una misura di randomness o di imprevedibilità: il suo valore sarà, ad esempio, maggiore per un serie di numeri completamente random rispetto al caso di un sistema periodico dove l'osservazione di un periodo rende predicibili i valori assunti in un periodo successivo.

2.2.1 L'entropia dei processi stocastici

Consideriamo una sorgente che emette una stringa S di simboli s_i tratti da un alfabeto $A = \{1, 2, \dots, b\}$, con una data probabilità $P(s)$, $\forall s \in S$.

L'indeterminazione associata ad ogni evento dipende, ovviamente, dalla probabilità che questo si verifichi: se $P(s) = 1$ per un certo evento, l'uscita dell'esperimento è completamente determinata, quindi l'entropia è minima; se invece $P(s) = 1/b$ per ogni possibile evento, l'indeterminazione è massima.

Ogni osservazione aggiunge l'informazione che è associata con l'apparire del simbolo corrente rimuovendo così l'indeterminazione preesistente.

L'informazione media su tutte le uscite $s_i \in A$ dell'esperimento E è misurata dall'*entropia di Shannon* così definita

$$H(E) = - \sum_{i=1}^b P(s_i) \ln P(s_i), \quad (2.1)$$

dove $P = \{P(1), \dots, P(b)\}$ è la distribuzione di probabilità per E.

L'entropia H gode di tre proprietà fondamentali:

1) $H = 0$ se e solo se uno dei simboli ha probabilità 1 (con la convenzione che $x \ln(x) = 0$ per $x=0$);

2) poiché $h(x) = -x \ln(x)$ è continua, non negativa, concava in $(0,1]$ risulta

$$H(E) = \sum_{i=1}^b h(p_i) \leq bh\left(\frac{1}{b} \sum_{i=1}^b p_i\right) = bh\left(\frac{1}{b}\right) = \ln(b), \quad (2.2)$$

per ogni vettore di probabilità $P = \{p_1, \dots, p_b\}$; il valore massimo dell'entropia $\ln(b)$ è proprio quello che si ha nel caso più indeterminato (cioè quando $p_i = \frac{1}{b}$, per $\forall i$);

3) l'informazione è additiva per eventi composti da eventi reciprocamente indipendenti infatti

$$\ln P(s_1 s_2 \dots s_n) = \ln \prod_{i=1}^n P(s_i) = \sum_{i=1}^n \ln P(s_i).$$

Si dimostra che la funzione $H(p_1, \dots, p_b)$ (con $\sum_{i=1}^b p_i = 1$) dell'equazione (2.1) è la sola funzione continua che soddisfa le tre condizioni suddette [4].

In generale la probabilità di ogni simbolo dipende dai simboli precedentemente emessi e dalle loro probabilità. Gli effetti della memoria possono essere valutati considerando n esperimenti successivi correlati $\{A_i\}$. L'entropia congiunta degli n esperimenti è così definita:

$$H_n = H(A_1, A_2, \dots, A_n) = - \sum_{S:|S|=n} P(S) \ln P(S), \quad (2.3)$$

dove la somma è intesa su tutte le possibili sequenze S di lunghezza n.

Nel caso $n=2$ risulta:

$$H(A_1, A_2) = - \sum_{i,j} P(s_i s_j) \ln P(s_i s_j) = \quad (2.4)$$

$$\begin{aligned}
&= - \sum_{i,j} P(s_i)P(s_j | s_i) \ln[P(s_i)P(s_j | s_i)] = \\
&= - \sum_i P(s_i) \ln P(s_i) - \sum_i P(s_i) \sum_j \ln[P(s_j | s_i) \ln P(s_j | s_i)],
\end{aligned}$$

dato che $\sum_j P(s_j | s_i) = 1, \forall i$.

Definendo

$$H(A_2 | s_i) = - \sum_j P(s_j | s_i) \ln P(s_j | s_i), \quad (2.5)$$

l'entropia dell'esperimento A_2 se A_1 ha prodotto s_i , la (2.5) diventa:

$$H(A_1, A_2) = H(A_1) + H(A_2 | A_1), \quad (2.6)$$

dove $H(A_2 | A_1) = \sum_i P(s_i)H(A_2 | s_i)$ è l'entropia condizionale di A_2 rispetto ad A_1 .

Si dimostra che

$$H(A_2 | A_1) \leq H(A_2), \quad (2.7)$$

cioè la conoscenza del risultato di A_1 riduce l'indeterminazione sul risultato di A_2 rispetto al caso di esperimenti completamente indipendenti. Risulta quindi:

$$H(A_1, A_2) \leq H(A_1) + H(A_2), \quad (2.8)$$

e per n generico

$$H_n = H(A_1, A_2, \dots, A_n) = \sum_{i=1}^n H(A_i | A_0, \dots, A_{i-1}) \leq H(A_1) + H(A_2) + \dots + H(A_n), \quad (2.9)$$

L'informazione media per simbolo $\frac{H_n}{n}$ quantifica la dipendenza delle sequenze di lunghezza n da quelle di lunghezza $(n-1)$.

L'entropia metrica di Kolmogorov è definita come

$$K_1 = \lim_{n \rightarrow \infty} \frac{H_n}{n} \quad (2.10)$$

e tale limite esiste per una sorgente stazionaria.

Dalla (2.9) segue che $H_n \leq n \ln b$ e quindi che l'entropia K_1 è compresa tra 0 e $\ln b$.

Si può verificare che

$$\lim_{n \rightarrow \infty} \frac{H_n}{n} = \lim_{n \rightarrow \infty} (H_{n+1} - H_n),$$

e dato che $H_{n+1} - H_n = H(A_{n+1} | A_0, \dots, A_n) \geq 0$ (cioè l'entropia congiunta cresce con n), l'entropia metrica rappresenta il tasso con cui l'entropia congiunta cresce con n [4].

I segnali random hanno un alto valore di entropia metrica, infatti l'emissione di un nuovo simbolo aumenta in maniera considerevole il contenuto di informazione a disposizione. Per un linguaggio con le sue regole grammaticali, invece, esiste già un limite alle possibili concatenazioni dei simboli dell'alfabeto, esiste quindi una certa predicibilità e il contenuto di informazione che aggiunge la conoscenza di un nuovo simbolo è ridotto rispetto al caso random.

2.2.2 Entropia dei sistemi dinamici

Nel paragrafo precedente abbiamo definito e interpretato le entropie di Shannon e Kolmogorov-Sinai per i processi stocastici.

Passiamo ora a considerare i sistemi dinamici cercando le definizioni operative delle entropie suddette per questi sistemi. Vedremo come attraverso una procedura di coarse-graining dello spazio delle fasi è possibile considerare il sistema dinamico come sorgente di un processo stazionario a stati finiti ¹.

Un sistema dinamico può in generale essere definito come una prescrizione matematica deterministica per l'evoluzione di uno stato nel tempo, dove quest'ultimo può essere una variabile continua o una variabile discreta.

Nel caso continuo, il sistema può essere descritto da un sistema di equazioni differen-

¹ sotto particolari ipotesi che consideremo

ziali²

$$\dot{x}_i(t) = f_i(x_1(t), \dots, x_n(t), t, \lambda)$$

con λ parametro di controllo. La conoscenza dello stato iniziale, di t , di λ determina univocamente lo stato all'istante t

$$x_i(t) = \psi(x_1(t_0), \dots, x_n(t_0), t, \lambda).$$

Nel caso discreto, l'evoluzione dello stato nello spazio delle fasi può essere descritta da una mappa

$$\vec{x}_{n+1} = F_\lambda(\vec{x}_n),$$

dove \vec{x}_n è il vettore nello spazio delle fasi che rappresenta lo stato del sistema all'istante n . Iterando l'operazione descritta dalla mappa per n volte si ottiene lo stato all'istante n a partire dallo stato \vec{x}_0 :

$$\vec{x}_n = F_\lambda^n(\vec{x}_0).$$

Nel caso continuo, si può ottenere una discretizzazione del tempo e quindi una mappa F considerando la sequenza dei punti intersezione del flusso con un'opportuna superficie di Poincarè.

Considereremo nel seguito sempre sistemi dinamici discreti.

Sia $X \in \mathbb{R}^d$ una regione limitata dello spazio delle fasi entro cui il moto è circoscritto, consideriamo una sua arbitraria partizione $\mathcal{B} = \{B_0, \dots, B_{b-1}\}$ con b componenti disgiunte ($\cup_{j=0}^{b-1} B_j = X$ e $B_j \cap B_k = \emptyset$ per $j \neq k$), ottenendo così una specie di griglia delle coordinate per la dinamica.

Sotto l'azione della mappa F , il sistema descrive un'orbita $\mathcal{O} = \{\vec{x}_0, \vec{x}_1, \dots, \vec{x}_n\}$ che tocca diversi elementi di \mathcal{B} . Sia A l'insieme delle etichette degli elementi della partizione $A = \{0, \dots, b-1\}$ e chiamiamo $s_i \in A$ l'indice del dominio B visitato dall'orbita

² ci sono casi in cui le equazioni differenziali sono di ordine superiore al primo

all'istante i . All'orbita \mathcal{O} è associata la sequenza $S = \{s_0, s_1, \dots, s_n\}$, dove $\vec{x}_i \in B_{s_i}$ per tutti gli $i = \{0, \dots, n\}$.

Con una particolare scelta di \mathcal{B} è possibile ricostruire la traiettoria originale \mathcal{O} , a partire dalla conoscenza della sequenza simbolica S , della partizione \mathcal{B} e della legge dinamica F . La sequenza S può infatti essere prodotta solo dai punti \vec{x}_0 appartenenti a

$$B_S = B_{s_0} \cap F^{-1}(B_{s_1}) \cap \dots \cap F^{-n}(B_{s_n}). \quad (2.11)$$

L'osservazione dei primi due simboli $\{s_0, s_1\}$ in S , per esempio, indica che il punto rappresentativo era all'istante $i = 0$ non solo in B_{s_0} ma anche in $F^{-1}(B_{s_1})$, cioè nella prima preimmagine di B_{s_1} . La misura congiunta agli istanti $i = 0$ e $i = 1$ riduce l'indeterminazione su \vec{x}_0 e le cose vanno sempre meglio considerando via via sequenze simboliche S più lunghe.

X è un set compatto dotato di una metrica ed è il supporto di una misura di probabilità μ invariante sotto la mappa F (cioè è invariante sotto l'evoluzione temporale). La probabilità $\mu(B_i)$ con $i \in A$ indica la frequenza con cui B_i viene visitato dall'orbita \mathcal{O} che descrive il sistema.

Possiamo definire l'*entropia di Shannon* per la partizione \mathcal{B} come

$$H(\mathcal{B}) = - \sum_{i=0}^{b-1} \mu(B_i) \ln \mu(B_i). \quad (2.12)$$

Costruiamo una successione di partizioni con la seguente procedura di "raffinazione". Formiamo a partire dalla partizione originale \mathcal{B} gli insiemi $F^{-1}(B_{s_i})$ e consideriamo per ogni coppia $s_j, s_k \in A$ le b^2 intersezioni

$$B_{s_j} \cap F^{-1}(B_{s_k}).$$

Tutte le intersezioni non vuote etichettate da $\{s_j, s_k\}$ costituiranno gli elementi della partizione \mathcal{B}_1 . Con le intersezioni non vuote del tipo

$$B_{s_j} \cap F^{-1}(B_{s_k}) \cap F^{-2}(B_{s_i})$$

formiamo la partizione \mathcal{B}_2 , e così via fino alla \mathcal{B}_n formata dalle intersezioni non vuote

$$B_{s_j} \cap F^{-1}(B_{s_k}) \cap F^{-2}(B_{s_i}) \cap \dots \cap F^{-(n-1)}(B_{s_n}).$$

Si nota che la successione delle partizioni è prodotta dalla dinamica stessa. La partizione \mathcal{B}_n è la partizione generata da \mathcal{B} in un intervallo di tempo di lunghezza n . Osserviamo che per una traiettoria $\mathcal{O} = \{\vec{x}_0, \vec{x}_1, \dots, \vec{x}_n\}$ gli eventi *elementari* sono del tipo $\vec{x}_i \in B_{s_i}$ e gli eventi *composti* di ordine n prendono la forma: $\vec{x}_{i+1} \in B_{s_{i+1}}, \dots, \vec{x}_{i+n} \in B_{s_{i+n}}$, che equivale alla condizione $\vec{x}_{i+1} \in B_{s_{i+1}} \cap \dots \cap F^{-n+1}(B_{s_{i+n}})$, cioè \vec{x}_{i+1} appartiene ad un elemento di \mathcal{B}_n .

Calcolando, mediante la (2.12), $H(\mathcal{B}_n)$ si ottiene l'entropia del blocco n dato che sommare su tutti gli elementi di \mathcal{B}_n significa sommare su tutte le possibili sequenze di lunghezza n .

Si dimostra l'esistenza del seguente limite

$$\begin{aligned} K_1(\mu, F, \mathcal{B}) &= \lim_{n \rightarrow \infty} \frac{H(\mathcal{B}_n)}{n} = \\ &= \lim_{n \rightarrow \infty} [H(\mathcal{B}_{n+1}) - H(\mathcal{B}_n)], \end{aligned} \quad (2.13)$$

per ogni partizione misurabile numerabile³. Questa quantità, detta entropia della trasformazione F rispetto alla partizione \mathcal{B} , rappresenta l'indeterminazione media per time-step dell'elemento di partizione che è correntemente visitato dall'orbita.

L'entropia metrica può essere pensata come una misura del tasso di creazione di informazione durante l'evoluzione temporale del sistema. Ci si può convincere del fatto che un sistema dinamico caotico crea informazione durante la sua evoluzione.

Data la sua sensibilità alle condizioni iniziali, orbite inizialmente vicine possono con il tempo divergere. In pratica noi possiamo distinguere la locazione delle orbite nello spazio delle fasi solo con un certo grado di accuratezza, è probabile, quindi, che due condizioni iniziali diverse possono apparire uguali entro questi limiti. Man mano che

³ Petersen, 1989

il sistema evolve le due orbite possono eventualmente distanziarsi tanto da diventare distinguibili. Supponendo di conoscere le equazioni dinamiche del sistema, possiamo usare l'osservazione ad istanti successivi per poter acquisire ulteriori informazioni sulle condizioni iniziali tanto da riuscire eventualmente a distinguerle.

La definizione dell'entropia metrica deve essere tale da produrre un tasso di creazione di informazione maggiore di zero per i sistemi caotici. Scegliendo in maniera inappropriata \mathcal{B} si rischia, invece, di avere un'entropia nulla anche se il moto è caotico; è necessario, quindi, considerare l'estremo superiore su tutte le possibili partizioni finite \mathcal{B} .

Definiamo così l'*entropia metrica* o di *Kolmogorov-Sinai* $K_1(\mu, F)$ per la trasformazione F come

$$K_1(\mu, F) = \sup_{\mathcal{B}} K_1(\mu, F, \mathcal{B}), \quad (2.14)$$

calcoliamo così l'informazione massima che si può ottenere da una misura eseguita sul sistema considerando tutte le possibili partizioni finite [4], [12].

L'entropia di Kolmogorov di un segnale sperimentale quantifica la caoticità di un segnale: un segnale regolare ha entropia $K_1 = 0$, un segnale puramente causale (random) ha $K_1 = \infty$. Un segnale deterministico caotico è caratterizzato da un valore finito positivo di K_1 che è correlato al "tempo medio di predicibilità" T dalla relazione $K_1 \sim T^{-1}$ dove T è il tempo medio nel futuro per cui la conoscenza dello stato del sistema può essere usata per predire la sua evoluzione [13].

La (2.14) con il calcolo dell'estremo superiore su tutte le partizioni sembra scoraggiare l'applicazione sperimentale; è comunque possibile evitare questo calcolo se si sceglie come partizione iniziale una partizione generatrice. Una partizione \mathcal{B} è detta generatrice se la partizione \mathcal{B}_n diventa infinitamente fine per $n \rightarrow \infty$; in questo caso la sua entropia rispetto al sistema dinamico è uguale all'entropia metrica cioè

$$K_1(\mu, F, \mathcal{B}) = K_1(\mu, F). \quad (2.15)$$

Ogni *refinement* di una partizione generatrice è anch'essa una partizione generatrice. In molti casi è più che plausibile che una partizione generatrice possa essere rifinita in una partizione uniforme con ε sufficientemente piccolo. Cohen e Procaccia indicano una formula per l'entropia metrica meno precisa ma sicuramente più semplice da interpretare rispetto alla (2.14) e che considera appunto una partizione uniforme.

Supponiamo di misurare lo stato del sistema, cioè la sua posizione nello spazio delle fasi, ad intervalli di tempo τ e di dividere lo spazio delle fasi in ipercubi di volume ε^d , dove d è il numero di gradi di libertà del sistema ed ε rappresenta l'accuratezza con cui ne misuriamo lo stato [13].

Sia $p(i_1, i_2, \dots, i_n)$ la probabilità congiunta che il punto rappresentativo dello stato del sistema all'istante $t = \tau$ sia nel volumetto i_1 , all'istante $t = 2\tau$ in i_2 e ... a $t = n\tau$ in i_n . Si può calcolare l'entropia metrica del sistema come

$$K = - \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n\tau} \sum_{i_1, \dots, i_n} p(i_1, \dots, i_n) \ln p(i_1, \dots, i_n). \quad (2.16)$$

2.2.3 Entropia delle serie temporali reali

Sono state sviluppate diverse formule che ci permettono di calcolare una stima dell'entropia di Kolmogorov-Sinai a partire dalla conoscenza della serie temporale. Illustriamo la formula sviluppata da Eckmann e Ruelle [16], che si rifa ai lavori precedenti di Grassberger e Procaccia e che sta alla base dello sviluppo di formule successive di facile implementazione in situazioni sperimentali ed è diventata una misura standard per l'entropia metrica.

Consideriamo una serie temporale di dati $\{u(1), u(2), \dots, u(N)\}$ acquisiti ad intervalli di tempo regolari τ .

Fissato un intero positivo m ed un numero reale positivo r , costruiamo i vettori

$\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N - m + 1)$ in \mathbb{R}^m così definiti:

$$\mathbf{x}(i) = [u(i), u(i + 1), \dots, u(i + m - 1)]. \quad (2.17)$$

Definiamo per ogni i , con $1 \leq i \leq N - m + 1$,

$$C_i^m(r) = \frac{1}{(N-m+1)} (\text{numero di vettori } \mathbf{x}(j) \text{ tali che } d[\mathbf{x}(i), \mathbf{x}(j)] \leq r).$$

Adottiamo per la distanza d tra i vettori la definizione di Takens:

$$d[\mathbf{x}(i), \mathbf{x}(j)] = \max_{k=1,2,\dots,m} |u(i+k-1) - u(j+k-1)|, \quad (2.18)$$

cioè la distanza tra due vettori è pari al massimo valore assoluto delle differenze tra le componenti corrispondenti. Possiamo interpretare $C_i^m(r)$ come la probabilità che il segnale $u(j+k)$ rimanga nell'intervallo di raggio r intorno a $u(i+k)$ per m unità di tempo consecutive. I valori $C_i^m(r)$ rappresentano una misura della frequenza di sequenze di lunghezza m simili a quella data entro un parametro di tolleranza r .

Definiamo

$$\Phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (2.19)$$

e l'entropia di Eckmann e Ruelle (E-K)

$$K_{E-R} = \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} [\Phi^{m+1}(r) - \Phi^m(r)] = \tau K_1(\mu) \quad (2.20)$$

Nota che

$$\Phi^{m+1}(r) - \Phi^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \ln C_i^{m+1}(r) - \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \simeq$$

(per N grande)

$$\simeq \frac{1}{N - m} \sum_{i=1}^{N-m} [\ln C_i^{m+1}(r) - \ln C_i^m(r)] = \frac{1}{N - m} \sum_{i=1}^{N-m} \ln \frac{C_i^{m+1}(r)}{C_i^m(r)}. \quad (2.21)$$

E' possibile, quindi, interpretare $\Phi^{m+1}(r) - \Phi^m(r)$ come la media su i del logaritmo della probabilità condizionale che

$|u(j+m) - u(i+m)| \leq r$ posto che $|u(j+k) - u(i+k)| \leq r$
per $k = 0, 1, \dots, m-1$.

Data la presenza dei limiti suddetti, la formula dell'entropia di E-R non risulta adatta all'analisi di serie temporali rumorose e di lunghezza finita N che è proprio il tipo di serie che si incontra nello studio di sistemi biologici.

Si deve a Pincus [17] l'introduzione di una famiglia di misure del sistema detta *entropia approssimata* così definita

$$ApEn(m, r) = \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)] \quad (2.22)$$

che risulta invece più facilmente applicabile alle serie sperimentali dato che tali parametri possono essere stimati dalla relativa famiglia di statistiche

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (2.23)$$

con N relativamente piccolo.

Risultano valori più bassi di $ApEn$ per le serie temporali più regolari mentre valori più alti per quelle meno predicibili (più complesse). L'entropia approssimata non va intesa comunque come valore approssimato dell'entropia di E-R, essa è piuttosto una statistica di regolarità. Ad esempio, un valore diverso da zero dell'entropia di E-R assicura che un sistema deterministico è caotico mentre l' $ApEn$ non può certificare il caos. L' $ApEn$, inoltre, è una famiglia di misure (dipende da m e da r), quindi nel confronto di due sistemi è necessario assumere gli stessi valori di m e di r .

Dati N punti, servono delle linee guida per scegliere dei valori di m ed r che assicurino delle stime ragionevoli di $ApEn(m, r)$ mediante $ApEn(m, r, N)$. Basandosi su calcoli che includono sia l'analisi teorica che le applicazioni cliniche, Pincus e Goldberger osservano che in molti casi per $m=2$ e $N=1000$, ad esempio, valori di r da 0.1 a 0.25 volte la deviazione standard delle serie forniscono una stima di $ApEn$ attraverso $ApEn(m, r, N)$ statisticamente valida. In generale essi suggeriscono un valore di N

almeno superiore a $10^m - 30^m$ per avere delle stime ragionevoli, e valori non troppo piccoli di r per evitare stime delle probabilità condizionate troppo povere nè valori troppo grandi per non rischiare di perdere dell'informazione più dettagliata. Inoltre affinché il parametro r funzioni da filtro per il rumore è necessario che esso sia più grande della maggior parte del rumore [14]. Va detto che, per evitare l'eventuale valore $\ln(0)$ e in accordo con l'entropia di Kolmogorov-Sinai, l'algoritmo $ApEn(m,r,N)$ conta in $C_i^m(r)$ anche la sequenza originale tra le sequenze vicine ad essa entro r . Richmann e Moorman [18] osservano che da questo *bias* (errore sistematico) risultano due rilevanti difetti dell'entropia approssimata:

- il valore atteso di $ApEn(m,r,N)$ cresce asintoticamente verso $ApEn(m,r)$ al crescere di N , quindi può avere solo un valore relativo utile nel confronto di serie di uguale lunghezza
- manca di consistenza relativa, cioè può capitare che l' $ApEn$ di un dato set risulti più alta di quella di un altro per alcuni valori di m , r e N ma più bassa per altri.

Quest'ultimo difetto è particolarmente grave se si pensa che l' $ApEn$ è usata spesso per confrontare due set di dati.

Per ridurre questo bias, Richmann e Moorman introdussero una nuova famiglia di misure detta *entropia di campionamento*⁴ $SE(m,r)$ così definita:

$$SE(m,r) = - \lim_{n \rightarrow \infty} \ln \frac{U^{m+1}(r)}{U^m(r)} \quad (2.24)$$

che è stimata dalla statistica

$$SE(m,r,N) = - \ln \frac{U^{m+1}(r)}{U^m(r)}. \quad (2.25)$$

⁴ il nome si riferisce alla sua applicabilità a serie temporali campionate da un processo continuo

dove gli $U^m(r)$, differiscono dai $C^m(r)$ solo perché non considerano il confronto delle sequenze con se stesse. Se si considera che l'entropia è concepita come il tasso di produzione di informazione, è chiaro che confrontare le sequenze con se stesse non ha senso e quindi $ApEn(m,r,N)$ sovrastima la regolarità della serie. Un'ulteriore differenza è che l' $ApEn$ considera le probabilità condizionate per vettori \mathbf{x}_i presi singolarmente e poi fa la media dei logaritmi di queste probabilità, la SE calcola la media delle probabilità condizionate e poi ne fa il logaritmo. Richmann e Moorman hanno dimostrato che la statistica SE

- risulta più in accordo con la teoria rispetto alla $ApEn$ per i numeri random con una distribuzione di probabilità nota,
- conserva la consistenza relativa in casi in cui l' $ApEn$ ne è priva,
- è meno sensibile alla lunghezza finita della serie rispetto all' $ApEn$.

In conclusione, la SE fornisce una più valida misura della regolarità di una serie temporale e ha trovato perciò vasta applicazione nello studio di segnali fisiologici.

2.2.4 L'entropia approssimata e le serie fisiologiche

Gli algoritmi appena visti $ApEn$ e SE sono stati applicati allo studio di segnali fisiologici come le serie di variabilità del ritmo cardiaco. Calcolando le entropie di campionamento di queste serie per certe patologie e confrontandole con le stesse per soggetti sani, risultano dei valori più alti per le serie patologiche. Potremmo interpretare questo valore più alto di entropia come indice di maggiore complessità. Questa conclusione sarebbe, tuttavia in contrasto con il dato già acquisito con altri metodi di una perdita di complessità per certe patologie e con l'invecchiamento. Un gran

numero di patologie sembrano degradare il contenuto d'informazione delle serie e ridurre la capacità di adattamento dell'individuo. Risultano, invece, delle uscite più random per i soggetti malati e delle correlazioni a lungo termine per i soggetti sani. Un possibile motivo per cui questi algoritmi tradizionali falliscono nel quantificare correttamente la complessità di queste serie è il fatto che sono misure su una singola scala temporale, mentre le serie temporali riferite a sistemi complessi possono presentare strutture su scale temporali multiple.

Nel 1991, Zhang, partendo da queste considerazioni, propone una misura di complessità che tiene conto delle scale multiple. Partendo dal segnale originale $x(t)$ costruisce dei segnali $y(t)$ dipendenti da τ che rappresentano le medie di $x(t)$ su intervalli di tempo τ : passando da $x(t)$ a $y(t)$ vengono eliminate le fluttuazioni su scala τ . Indicando con $S(\tau)$ l'entropia di Shannon calcolata per la serie $y(t)$, Zhang definisce la complessità come l'integrale delle entropie riferite a tutte le possibili scale:

$$K = \int_1^N d\tau S(\tau).$$

Questa misura ha il pregio di attribuire una maggiore complessità al rumore correlato $\frac{1}{f}$ rispetto al rumore non correlato (rumore bianco).

2.2.5 Il metodo MSE

Costa, Goldberger e Peng osservano la necessità di sostituire nel metodo di Zhang al calcolo delle entropie di Shannon quello delle entropie di campionamento quando si passa a serie temporali sperimentali. Descriviamo brevemente il metodo da loro introdotto noto come analisi multiscala dell'entropia MSE.

Data una serie temporale x_1, \dots, x_N , costruiamo la serie temporale coarse-grained,

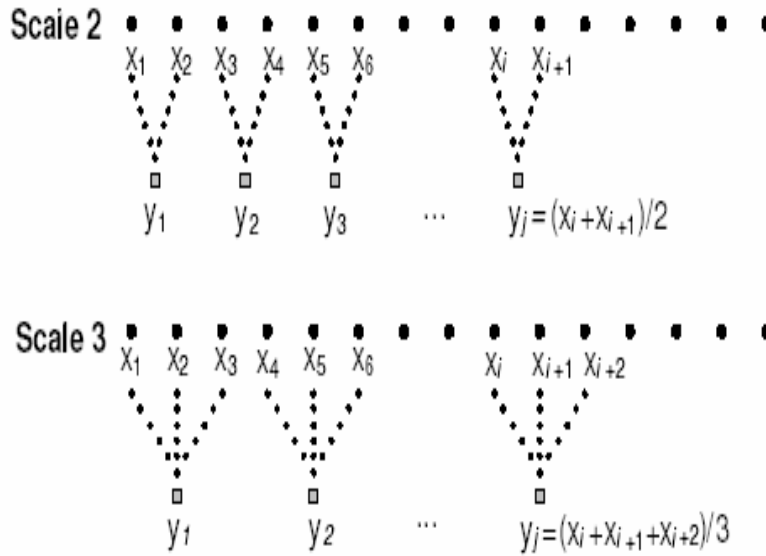


Figura 2.1: Rappresentazione schematica della procedura di coarse-graining [8]

secondo il fattore di scala τ , $\{y_j^{(\tau)}\}$ così definita:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i \text{ con } 1 \leq j \leq \frac{N}{\tau}. \quad (2.26)$$

Osserviamo che il fattore di scala specifica il numero di valori dalla cui media è stato ottenuto ogni elemento delle serie coarse-grained. Per ognuna delle serie $y_j^{(\tau)}$ si deve calcolare l'entropia di campionamento SE .

E' utile, a questo punto, per confrontare due segnali in base alla loro complessità diagrammare le diverse SE rispetto ai fattori di scala τ relativi.

La validità del metodo è stata ancora una volta testata sulle serie di rumore bianco e di rumore $\frac{1}{f}$: l'entropia del rumore bianco decresce monotonamente con il fattore di scala mentre quella riferita al rumore rosa $\frac{1}{f}$ rimane all'incirca costante. Questi risultati sono consistenti con il fatto che il rumore $\frac{1}{f}$ ha una struttura su scale temporali multiple, è cioè più complesso del rumore bianco [8].

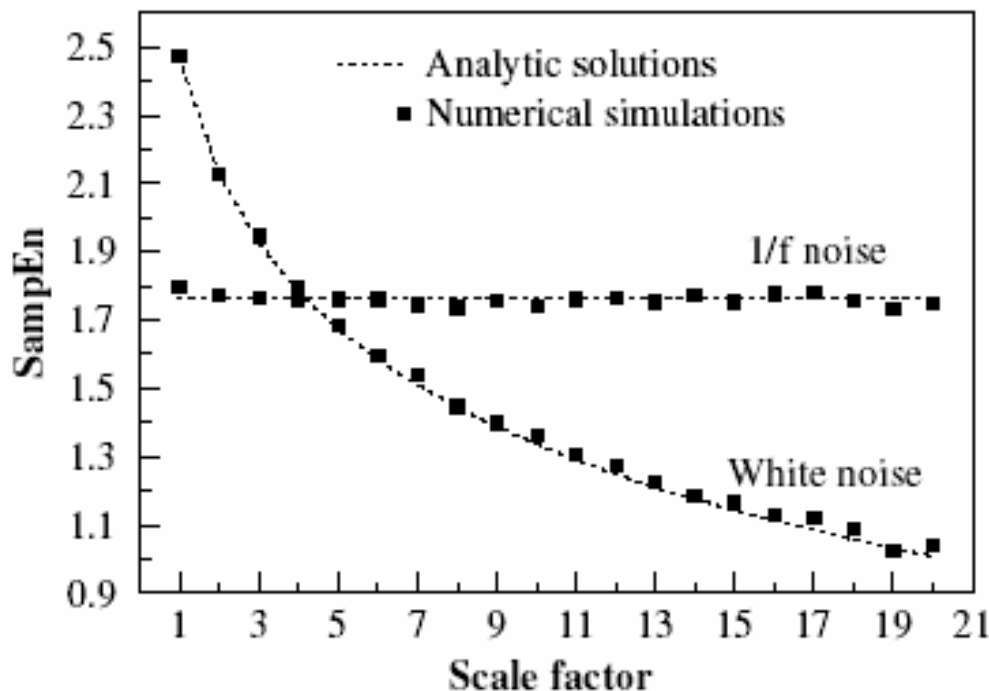


Figura 2.2: Analisi MSE per il rumore bianco gaussiano e per il rumore $1/f$ [8].

Gli autori hanno applicato tale metodo a serie RRI ottenute da registrazioni elettrocardiografiche di 24 ore [15].

Il metodo MSE applicato alle nostre serie RRI che sono, invece, delle serie short-term conferma la teoria generale della perdita di complessità nei malati a tutte le scale (in misura maggiore in regime di respirazione basale). L'analisi MSE delle serie SAP e DAP mostra un perdita di complessità nei malati solo a grandi valori del fattore di scala τ . Dall'analisi delle serie ILV risulta che l'entropia è più alta nei sani solo a scale temporali piccole in regime di respirazione basale mentre è sempre inferiore a quella dei malati in regime di respirazione controllato. Emerge un'influenza diversa della respirazione controllata nelle due classi: la respirazione controllata è meno efficace nel regolarizzare le serie ILV nei malati. I valori medi delle diverse SE e i loro errori standard sono riportati nelle tabelle in appendice. (Precisiamo che è stato utilizzato

un valore di r pari allo 0.15% della deviazione standard di ciascuna delle serie riscalate e m pari ad uno).

Osserviamo che le nostre serie, dopo il pretrattamento, sono serie di circa 600 punti (5 minuti di registrazione con un tasso di campionamento di 2Hz) quindi sono stati scelti fattori di scala $\tau \leq 10$ per aver una statistica sufficiente anche alle scale più grandi (60 punti per le serie a $\tau = 10$). Questo implica che la nostra analisi è limitata ad una parte della banda ad alte frequenze HF.

La banda da noi considerata comprende la frequenza respiratoria 0.25 Hz e ricordiamo che a questa frequenza la respirazione influenza direttamente sia le variazioni di RRI che quelle di SAP e DAP. Mentre non abbiamo potuto analizzare l'influenza che la respirazione esercita sulle serie RRI, SAP e DAP indirettamente attraverso il loop di controllo, ricordiamo, infatti, che questa ha un picco intorno alla frequenza di 0.1 Hz.

2.3 Modello autoregressivo multivariato:VAR

Dopo aver stimato l'imprevedibilità delle serie alle diverse scale, si può pensare ad un modello del sistema che tenga conto delle sorgenti di imprevedibilità e dello scaling temporale. Abbiamo già detto che la complessità delle fluttuazioni suggerisce tra le sorgenti di imprevedibilità la dinamica caotica oltre ad una componente di rumore esterno, ma un modello che tenga conto in maniera esatta delle due componenti non è applicabile [5]. E' possibile assumere che la non linearità sia una piccola perturbazione di un sistema essenzialmente stocastico lineare, o, alternativamente, che l'elemento di rumore esterno sia una piccola perturbazione di un processo deterministico non lineare. Seguendo il primo approccio e tenendo conto dell'accoppiamento tra le diverse

serie di variabilità cardiovascolare tramite i diversi meccanismi di regolazione è stato costruito un modello autoregressivo vettoriale VAR alle diverse scale temporali delle nostre serie cardiovascolari.

Come vedremo nel prossimo paragrafo, il modello VAR assunto, essendo time-invariant, assume la stazionarietà dei dati, cioè che le proprietà statistiche del sistema (medie, varianze, etc) non cambino nel tempo. E' noto che, sotto condizioni stabili, la massima durata nel sistema cardiovascolare di una registrazione stazionaria è tipicamente intorno ai dieci minuti. In questo studio l'ipotesi di stazionarietà è, quindi, giustificata dato che le nostre registrazioni in regime basale e controllato sono entrambe rispettivamente di circa otto minuti.

Nel seguito esponiamo un'introduzione formale al modello autoregressivo univariato AR e a quello vettoriale VAR.

2.3.1 Il modello AR

Poichè le equazioni deterministiche lineari del moto con coefficienti costanti possono avere solo soluzioni esponenziali o periodiche (forse anche con più di un periodo), i sistemi lineari hanno sempre bisogno di input irregolari esterni per produrre segnali irregolari. Il più semplice sistema che produce segnali non periodici è un processo stocastico lineare. Un processo stocastico (con valori in un insieme A) è una collezione di elementi aleatori (con valori in A), muniti di un indice t che varia in un insieme dato T . I valori che le variabili possono assumere caratterizzano gli stati del processo. Una classe di processi stocastici frequentemente discussa nell'analisi di serie temporali è quella dei processi autoregressivi.

Siano $X(t)$ le variabili random agli istanti discreti $t=\{\dots,-1,0,1,\dots\}$. Si parla di processo autoregressivo di ordine p , denotato con $AR(p)$, quando le variabili random

soddisfano la seguente equazione differenziale:

$$X(t) = \sum_{k=1}^p \alpha_k X(t-k) + \eta(t) \quad (3.27)$$

dove $\eta(t) \propto WN(0, \sigma^2)$,

intendendo con $WN(0, \sigma^2)$ un segnale di rumore gaussiano bianco con valor medio zero e varianza σ^2 . Indicando con $E(\eta(t))$ il valore di aspettazione della t-esima osservazione del segnale η , possiamo scrivere le seguenti relazioni:

$$\begin{aligned} E(\eta(t)) &= 0 \\ E(\eta^2(t)) &= \sigma^2 \\ E(\eta(t), \eta(\tau)) &= 0 \text{ per } t \neq \tau \end{aligned} \quad (3.28)$$

cioè la correlazione tra η ad istanti diversi è nulla (la correlazione si calcola infatti come $Corr(\eta(t), \eta(t-j)) = \frac{Cov(\eta(t), \eta(t-j))}{\sigma^2}$) [19]. Nota che la probabilità di una realizzazione X all'istante t è quella di una variabile random gaussiana con valore di aspettazione

$$E(X(t)) = \sum_{k=1}^p \alpha_k X(t-k). \quad (3.29)$$

E' interessante chiarire il legame tra questo particolare modello stocastico lineare e la serie temporale $X(t)$ di lunghezza N . Innanzitutto la serie temporale viene considerata come un campione di una realizzazione di un processo stocastico a tempo discreto e supponiamo di voler identificare un modello dinamico per descriverla. Cerchiamo quindi di esprimere $X(t)$ in funzione di $X(t-k)$, il valore di X agli istanti precedenti, con $k = 1, \dots, p$ (con p da determinare). Scegliamo la relazione più semplice cioè $X(t)$ è combinazione lineare dei valori $X(t-k)$:

$$X(t) = \sum_{k=1}^p \alpha_k X(t-k). \quad (3.30)$$

Se esistesse un valore di p e di $\alpha = \{\alpha_1, \dots, \alpha_p\}$ per cui la relazione (3.30) risultasse verificata per $t = \{1, 2, \dots, N\}$ e fosse inoltre possibile assumere la stazionarietà del

processo, potremmo usare la relazione (3.30) per descrivere il processo anche in istanti t non appartenenti all'intervallo $[1, N]$. Avremmo così ottenuto il modello cercato. Di fatto è molto raro che ciò accada. Per questo conviene assumere che la (3.30) debba essere verificata a meno di un errore $\eta(t)$, chiedendo cioè che sia verificata l'equazione:

$$X(t) = \sum_{k=1}^p \alpha_k X(t-k) + \eta(t).$$

Essendo ora il numero delle incognite (p componenti di α e N valori di η) superiore al numero delle relazioni la soluzione esiste per ogni valore di p . È interessante individuare, se esiste, la soluzione per cui $\eta(t)$ è un rumore bianco perchè, in questo caso, in qualsiasi istante t la sola conoscenza dei p valori $X(t-k)$ è sufficiente per stimare nel modo migliore possibile $X(t)$ dato che lo scarto $\eta(t)$ non è ulteriormente spiegabile (algoritmicamente incompressibile). Si richiede inoltre che sia il rumore che $X(t)$ abbiano distribuzioni gaussiane perchè questo garantisce che $X(t)$ abbia la stessa distribuzione di $X(t-k)$ dato che la somma di distribuzioni gaussiane è gaussiana. Il modello individuato è appunto il modello AR(p) in (3.27).

2.3.2 Il modello VAR

In un processo AR (univariato) i valori correnti di una serie temporale sono legati ai valori precedenti della stessa serie. Spesso nella modellizzazione empirica si ottengono risultati migliori includendo altre variabili esplicative, ad esempio la pressione arteriosa sistolica SAP sull'intervallo RR. Il modello vettoriale autoregressivo VAR contiene non soltanto un modello di ciascuna serie ma anche un modello delle relazioni tra le variabili incluse.

Un processo VAR per N variabili di ordine p è definito dalle equazioni:

$$\begin{aligned}
 X_1(t) &= \sum_{k=1}^p \alpha_{11}(k)X_1(t-k) + \sum_{k=1}^p \alpha_{12}(k)X_2(t-k) + \dots + \sum_{k=1}^p \alpha_{1N}(k)X_N(t-k) + \eta_1(t) \\
 X_2(t) &= \sum_{k=1}^p \alpha_{21}(k)X_1(t-k) + \sum_{k=1}^p \alpha_{22}(k)X_2(t-k) + \dots + \sum_{k=1}^p \alpha_{2N}(k)X_N(t-k) + \eta_2(t) \\
 &\vdots \\
 X_N(t) &= \sum_{k=1}^p \alpha_{N1}(k)X_1(t-k) + \sum_{k=1}^p \alpha_{N2}(k)X_2(t-k) + \dots + \sum_{k=1}^p \alpha_{NN}(k)X_N(t-k) + \eta_N(t)
 \end{aligned} \tag{3.31}$$

oppure in forma matriciale:

$$\begin{bmatrix} X_1(t) \\ \vdots \\ X_N(t) \end{bmatrix} = \sum_{k=1}^p \begin{bmatrix} \alpha_{11}(k) & \dots & \alpha_{1N}(k) \\ \vdots & \ddots & \vdots \\ \alpha_{N1}(k) & \dots & \alpha_{NN}(k) \end{bmatrix} \begin{bmatrix} X_1(t-k) \\ \vdots \\ X_N(t-k) \end{bmatrix} + \begin{bmatrix} \eta_1(t) \\ \vdots \\ \eta_N(t) \end{bmatrix}$$

\Leftrightarrow

$$\mathbf{X}(t) = \sum_{k=1}^p \boldsymbol{\alpha}_k \mathbf{X}(t-k) + \boldsymbol{\eta}(t), \tag{3.32}$$

dove il vettore ($N \times 1$) $\boldsymbol{\eta}(t)$ è una generalizzazione vettoriale del rumore bianco gaussiano:

$$E(\boldsymbol{\eta}(t)) = \mathbf{0} \tag{3.33}$$

$$E(\boldsymbol{\eta}^2(t)) = \boldsymbol{\Omega}$$

$$E(\boldsymbol{\eta}(t), \boldsymbol{\eta}(\tau)) = \mathbf{0} \text{ per } t \neq \tau,$$

con $\boldsymbol{\Omega}$ una matrice ($N \times N$) simmetrica definita positiva (la matrice di varianze e covarianze degli $\boldsymbol{\eta}(t)$).

Lo stimatore mediante i minimi quadrati ordinari OLS (ordinary least squares) equazione per equazione produce stime consistenti e asintoticamente normali di tutti i

coefficienti delle matrici α_k ([19]). Ponendo

$$\mathbf{Y}(t) \equiv \begin{bmatrix} \mathbf{X}(t-1) \\ \mathbf{X}(t-2) \\ \vdots \\ \mathbf{X}(t-p) \end{bmatrix},$$

e la matrice $N \times (Np+1)$

$$\mathbf{\Pi} \equiv \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_p \end{bmatrix},$$

la stima di $\mathbf{\Pi}$ è data:

$$\hat{\mathbf{\Pi}} = \left[\sum_{t=1}^T \mathbf{X}(t) \mathbf{Y}'(t) \right] \left[\sum_{t=1}^T \mathbf{Y}(t) \mathbf{Y}'(t) \right]^{-1},$$

dove la riga j -esima di $\hat{\mathbf{\Pi}}$ è:

$$\hat{\pi}'_j = \left[\sum_{t=1}^T X_j(t) \mathbf{Y}'(t) \right] \left[\sum_{t=1}^T \mathbf{Y}(t) \mathbf{Y}'(t) \right]^{-1},$$

cioè il vettore dei coefficienti stimati da una regressione OLS di $X_j(t)$ su $\mathbf{Y}(t)$ [19].

2.4 L'interpretazione dei coefficienti

Riconsideriamo dapprima il processo univariato AR(1):

$$X(t) = \alpha X(t-1) + \eta(t)$$

dove con $\eta(t)$ indichiamo un processo casuale a valor medio nullo con varianza

$\sigma_\eta^2 = E(\eta(t)^2)$ formato da campioni indipendenti tra loro e da $X(t)$

($E((X(t)\eta(t)) = 0)$). Dato che il processo $X(t)$ è stazionario e a media nulla, avremo:

$$E(X^2(t)) = E(X^2(t-1)) = \sigma_X^2. \quad (4.34)$$

Otteniamo quindi il seguente legame tra σ_X^2 e σ_η^2 :

$$E((\alpha X(t-1) + \eta(t))^2) = \alpha^2 E(X^2(t)) + E(\eta^2(t)) = E(X^2(t))$$

da cui risulta

$$\sigma_\eta^2 = \sigma_X^2(1 - \alpha^2).$$

Quindi, quanto più α è vicino a 1, tanto più σ_η^2 è piccolo e tanto meno $X(t)$ si discosterà da $X(t-1)$. (Si può vedere facilmente che il coefficiente di correlazione con ritardo unitario $C(1) = \frac{E(X(t)X(t-1))}{E(X(t)^2)}$ è proprio uguale ad α [20]). Segue che tanto più α è prossimo ad 1 tanto meglio il valore che assumerà il processo casuale al tempo t è predicibile una volta noto il valore assunto all'istante $(t-1)$.

La miglior stima lineare del valore $X(t)$, una volta noto il valore assunto da $X(t-1)$, si riduce alla stima di α . Il metodo OLS, applicato alla regressione $X(t) = rX(t-1)$ ne offre una stima consistente.

Non dimentichiamo che con questo modello abbiamo valutato solo la predizione lineare. Il modello divide, infatti, la serie temporale in due componenti: la serie temporale predicibile e la sequenza dell'errore di predizione; in quest'ultima, la serie di rumore del segnale considerato, sono inclusi eventuali contributi non lineari.

Il modello VAR ha delle potenzialità maggiori: per ogni serie X_i considerata, esso permette di distinguere e valutare in tutto quello che abbiamo chiamato "rumore" il contributo lineare di alcuni processi che riteniamo siano connessi a $X_i(t)$: più questo contributo è piccolo meno cambia il valore di $X_i(t)$ rispetto a quello di $X_i(t-1)$, quindi più tale serie è predicibile.

In particolare consideriamo una delle equazioni in (3.31) nel caso $p=1$:

$$X_i(t) = \alpha_{ii}X_i(t-1) + \sum_{\substack{j=1:N \\ j \neq i}} \alpha_{ij}X_j(t-1) + \eta_i(t). \quad (4.35)$$

Procediamo analogamente al caso AR(1).

Considerando dei processi $X_j(t)$ con $j = 1 : \dots : N$ casuali stazionari a media nulla e

varianza unitaria possiamo scrivere:

$$E(X_j(t)^2) = E(X_j(t-1)^2) = \sigma_{X_j}^2$$

inoltre assumendo, oltre alle (3.33),

$$E(X_i(t)X_j(t)) = 0$$

$$E(X_j(t)\eta(t)) = 0$$

risulta:

$$\begin{aligned} E((\alpha_{ii}X_i(t-1) + \sum_{\substack{j=1:N \\ j \neq i}} \alpha_{ij}X_j(t-1) + \eta_i(t))^2) &= \alpha_{ii}^2 E(X_i(t-1)^2) + \\ &+ \sum_{\substack{j=1:N \\ j \neq i}} \alpha_{ij}^2 E(X_j(t-1)^2) + E(\eta_i(t)^2) \end{aligned}$$

da cui ricaviamo:

$$\sigma_{X_i}^2 = \alpha_{ii}^2 \sigma_{X_i}^2 + \sum_{\substack{j=1:N \\ j \neq i}} \alpha_{ij}^2 \sigma_{X_j}^2 + \sigma_{\eta_i}^2,$$

e ancora

$$1 = \alpha_{ii}^2 + \sum_{\substack{j=1:N \\ j \neq i}} \alpha_{ij}^2 + \sigma_{\eta_i}^2.$$

Ne segue che quanto più $(\sum_{j=1:N} \alpha_{ij}^2 + \sigma_{\eta_i}^2)$ è prossimo a 0, e quindi α_{ii}^2 è prossimo a 1, tanto meno il valore di $X_i(t)$ si discosterà da quello di $X_i(t-1)$ (come si vede dalla (4.35)), cioè il segnale risulta più predicibile.

Se il valore di α_{ij} è piccolo significa che X_j contribuisce poco (almeno linearmente) a far variare il segnale X_i .

I valori medi dei coefficienti del modello VAR applicato alle nostre serie sono riportati nelle tabelle in appendice C. Per quanto concerne la relazione causale di RRI su SAP risulta che in entrambi i regimi di respirazione il relativo coefficiente è negativo indicando una prevalenza del meccanismo di Windkessel che tende ad abbassare

la pressione all'aumentare dell'intervallo RRI. Questo coefficiente è in modulo più grande per i sani. Lo stesso si trova per il coefficiente dell'influenza di RRI su DAP. Il coefficiente di DAP su RRI, che fornisce indicazioni sull'attività dei baroriflessi, indica una forte discriminazione tra sani e malati: in condizioni di respiro basale il coefficiente è positivo per i sani e negativo per i malati. Inoltre nei malati esso è più influenzato dal respiro controllato, infatti è quasi zero nei malati e ancora positivo nei sani. L'interazione RRI su ILV è più forte nei sani nel caso di respiro controllato e $\tau = 4$. Troviamo anche che l'interazione ILV su RRI è positiva e più forte nei sani, in condizioni basali e ad alte frequenze ($\tau \leq 4$). Osserviamo per la serie RRI un coefficiente di autocorrelazione lineare maggiore per i malati in respiro basale e solo a grandi τ in respiro controllato. Il coefficiente di autocorrelazione lineare della serie DAP è sempre maggiore nei sani rispetto ai malati in regime di respiro controllato, e solo a piccoli τ in respiro basale. Troviamo significative differenze nei coefficienti dell'influenza di SAP su DAP: il coefficiente è sempre negativo per i sani e il suo modulo cresce con τ . La respirazione interagisce in un loop aperto: cioè non c'è un'influenza significativa di SAP su ILV, invece, c'è un'influenza di ILV su SAP che è maggiore nei malati.

Capitolo 3

La classificazione automatica e la selezione di variabili

3.1 La classificazione delle osservazioni

In questo capitolo dopo aver esposto brevemente la teoria della classificazione delle osservazioni esamineremo il metodo di riconoscimento automatico da noi utilizzato. L'obiettivo della classificazione automatica (pattern recognition) è dividere dei soggetti in diverse classi in base ai valori di alcune variabili su di essi misurate: formalmente, dato un set di n oggetti caratterizzati da p caratteristiche (feature), compito della classificazione è associare ad ogni oggetto una particolare caratteristica, l'etichetta di classe, come funzione delle p caratteristiche date. Questa funzione è detta funzione di predizione e per determinarla noi dobbiamo "imparare" dall'analisi dei valori assunti da queste caratteristiche nel caso di soggetti etichettati. Il modello più generale consiste di tre elementi:

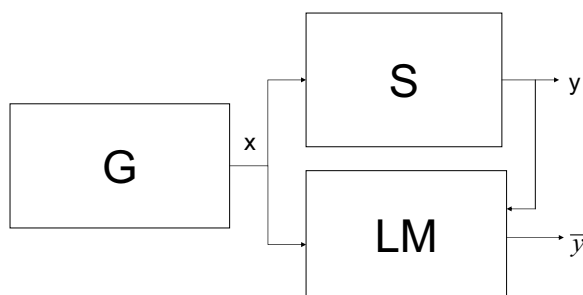


Figura 1.1: Un modello di learning machine

- il generatore G dei dati (gli esempi etichettati),
- l'operatore supervisore S ,
- la macchina che impara, *learning machine* LM .

Il generatore G genera i vettori $\mathbf{x} \in X$ indipendentemente e identicamente distribuiti secondo una funzione di distribuzione di probabilità $F(\mathbf{x})$ sconosciuta. L'operatore S agisce sui vettori \mathbf{x} restituendo in uscita i valori y , le rispettive etichette.

La LM osserva le n coppie $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, il *training set*, e costruisce un operatore che può essere usato per predire la risposta di S per un nuovo soggetto non etichettato. Più precisamente, la LM costruisce un'approssimazione dell'operatore supervisore, cioè sceglie in un fissato set di funzioni che essa può implementare quella che restituisce la migliore predizione per le uscite y [21].

Esistono diversi criteri per scegliere la funzione migliore ed altrettanti metodi di classificazione.

Il nostro è un problema di classificazione binaria: dobbiamo trovare un criterio per decidere circa l'appartenenza di un individuo alla categoria dei sani o dei malati sulla

base di un insieme di osservazioni, x_1, x_2, \dots, x_p , ad esso relative. Le nostre variabili sono le entropie delle serie RRI, DAP, SAP e ILV insieme ai coefficienti del modello VAR relativo alle stesse serie alle varie scale temporali $\tau = \{1, 2, \dots, 10\}$ e nei due diversi regimi di respirazione, basale e controllata: in totale abbiamo 400 variabili misurate su ognuno dei 303 soggetti.

Descriviamo nel paragrafo seguente l'analisi discriminante lineare di Fisher su cui si fonda il nostro metodo di classificazione.

3.2 Funzione discriminante lineare di Fisher

Sia $\mathbf{x} = \{x_1, x_2, \dots, x_p\}$ un vettore di variabili aleatorie con distribuzione di probabilità $f_1(\mathbf{x})$ in π_1 e $f_2(\mathbf{x})$ in π_2 , dove π_1 e π_2 sono le due popolazioni coinvolte nel problema. Una regola di classificazione può essere definita da una partizione disgiunta dello spazio dei campioni R in due regioni R_1 e R_2 insieme alla regola di decisione che assegna a π_1 individui che cadono in R_1 , a π_2 quelli in R_2 . Una regola intuitiva per definire l'appartenenza di un individuo ad una popolazione è la seguente:

$$\begin{aligned} \mathbf{x} \in \pi_1 & \text{ se } f_1(\mathbf{x}) \geq f_2(\mathbf{x}), \\ \mathbf{x} \in \pi_2 & \text{ se } f_1(\mathbf{x}) < f_2(\mathbf{x}), \end{aligned} \tag{2.1}$$

quindi in termini di R_1 e R_2 :

$$\begin{aligned} R_1 &= \{\mathbf{x} | f_1(\mathbf{x}) \geq f_2(\mathbf{x})\} \\ R_2 &= \{\mathbf{x} | f_1(\mathbf{x}) < f_2(\mathbf{x})\}. \end{aligned} \tag{2.2}$$

cioè colloco l'individuo nella popolazione a cui corrisponde la maggiore probabilità di appartenenza.

Stiamo in questo modo, tuttavia, trascurando due fattori:

- l'eventuale diversità delle probabilità a priori di osservare individui dalle due popolazioni (cioè, nel caso di classificazione tra sani e malati, l'incidenza della malattia)
- possibili diversi costi nel caso di errori di classificazioni nei due gruppi (considerando di nuovo le diagnosi mediche, classificare un paziente sano come malato può essere, per esempio, meno grave di classificarne uno malato come sano).

Ad esempio, se per un vettore \mathbf{x} risulta $f_1(\mathbf{x})$ leggermente più grande di $f_2(\mathbf{x})$ ma la probabilità a priori q_2 che un individuo appartenga a π_2 è maggiore della probabilità q_1 relativa a π_1 , potrei decidere contro la 2.1 di collocare l'individuo nella popolazione π_2 . Inoltre potrei pesare in maniera diversa la possibilità di sbagliare collocando un individuo di π_2 in π_1 ($C(1|2)$) rispetto all' errore opposto ($C(2|1)$).

si dimostra che regola ottima risulta:

$$R_1 = \left\{ \mathbf{x} \mid \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq k \right\}, \quad (2.3)$$

$$R_2 = \left\{ \mathbf{x} \mid \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < k \right\}$$

dove $k = \frac{C(1|2)q_2}{C(2|1)q_1}$ [22]. Nota che la 2.3 si riduce alla 4.22 per $C(1|2) = C(2|1)$ e $q_1 = q_2$ [22].

La (2.3) sembrerebbe, già, fornire la soluzione, ma essa in realtà contiene delle quantità ignote che vanno stimate dai dati forniti al calcolatore in fase di addestramento. Il problema della stima delle distribuzioni di probabilità è ancora oggetto di studio nella sua formulazione generale.

Noi ci occuperemo del caso più studiato in letteratura in cui le popolazioni seguono una distribuzione multinormale.

Supponiamo, in particolare, che $f_1(\mathbf{x})$ e $f_2(\mathbf{x})$ siano distribuzioni gaussiane multivariate con vettori delle medie rispettivamente $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ e con la stessa matrice di

dispersione Σ , cioè per ($i = 1, 2$)

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{1}{2}p} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)\right].$$

Ricordiamo che la matrice di dispersione è definita come:

$$\Sigma = E((\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu}')), \quad (2.4)$$

quindi i termini sulla diagonale rappresentano le varianze delle rispettive variabili ($\sigma_{ii} = \text{Var}(x_i)$) mentre il termine generico σ_{ij} è la covarianza tra le variabili x_i e x_j .

Il rapporto delle probabilità risulta quindi:

$$\begin{aligned} \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} &= \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_2)\right] = \\ &= \exp\left[-\frac{1}{2}\boldsymbol{\mu}'_1 \Sigma^{-1} \boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}'_2 \Sigma^{-1} \boldsymbol{\mu}_2 + \mathbf{x}' \Sigma^{-1} \boldsymbol{\mu}_1 - \mathbf{x}' \Sigma^{-1} \boldsymbol{\mu}_2\right]. \end{aligned} \quad (2.5)$$

Sostituendo questo rapporto nella 2.1 e passando ai logaritmi:

$$R_1 : -\frac{1}{2}\boldsymbol{\mu}'_1 \Sigma^{-1} \boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}'_2 \Sigma^{-1} \boldsymbol{\mu}_2 + \mathbf{x}' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \geq \ln(k). \quad (2.6)$$

Aggiungendo e sottraendo al membro sinistro $\frac{1}{2}\boldsymbol{\mu}'_2 \Sigma^{-1} \boldsymbol{\mu}_1$ e riarrangiando risulta:

$$R_1 : \mathbf{x}' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \geq \ln(k), \quad (2.7)$$

dove il primo termine del membro sinistro è la nota *funzione lineare discriminante* di Fisher. Essa è una funzione lineare delle componenti del vettore delle osservazioni. (Osserviamo che, nel caso di matrici di dispersione diverse nelle due popolazioni, saremmo giunti ad una funzione discriminante quadratica).

Supponendo ugualmente probabili le due popolazioni e uguali i costi degli errori di classificazione, assoceremo un individuo alle popolazioni 1 e 2 in base alla regola seguente:

$$\begin{aligned} R_1 &: \mathbf{x}' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \geq \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ R_2 &: \mathbf{x}' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) < \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). \end{aligned} \quad (2.8)$$

L'equazione

$$\mathbf{x}'\mathbf{w} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)'\mathbf{w}, \quad (2.9)$$

con $\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, rappresenta l'equazione dell'iperpiano nello spazio R^p dei vettori \mathbf{x} che separa le due popolazioni, il *decision boundary*.

Supponiamo di avere due campioni random e indipendenti $\{\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \dots, \mathbf{x}_{n_1}^{(1)}\}$ e $\{\mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)}, \dots, \mathbf{x}_{n_2}^{(2)}\}$ tratti dalle due popolazioni p-variate di cui non si conoscono esattamente le medie e le matrici di dispersione. Supponendo le due popolazioni normali, stimiamo queste quantità a partire dai campioni come segue:

$$\begin{aligned} \bar{\mathbf{x}}^{(1)} &= \sum_{\alpha=1}^{n_1} \frac{\mathbf{x}_\alpha^{(1)}}{n_1}, \\ \bar{\mathbf{x}}^{(2)} &= \sum_{\alpha=1}^{n_2} \frac{\mathbf{x}_\alpha^{(2)}}{n_2}, \\ S &= \frac{1}{n_1 + n_2 - 2} \left[\sum_{\alpha=1}^{n_1} (\mathbf{x}_\alpha^{(1)} - \bar{\mathbf{x}}^{(1)})(\mathbf{x}_\alpha^{(1)} - \bar{\mathbf{x}}^{(1)})' + \sum_{\alpha=1}^{n_2} (\mathbf{x}_\alpha^{(2)} - \bar{\mathbf{x}}^{(2)})(\mathbf{x}_\alpha^{(2)} - \bar{\mathbf{x}}^{(2)})' \right]. \end{aligned} \quad (2.10)$$

La stima S , detta matrice di covarianza campionaria comune (pooled sample covariance matrix), si dimostra essere uno stimatore non distorto di $\boldsymbol{\Sigma}$ nel caso di matrici di dispersione delle due popolazioni uguali $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. Condizione necessaria per la definita positività della matrice S di è $n_1 + n_2 \geq p + 1$.

Se le osservazioni \mathbf{x}_i sono effettivamente indipendenti, comunque, la condizione è anche sufficiente.

A questo punto sostituendo le stime campionarie nella 2.8 risulta:

$$\begin{aligned} R_1 &: \mathbf{x}'\mathbf{w} \geq \frac{1}{2}(\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)})'\mathbf{w} \\ R_2 &: \mathbf{x}'\mathbf{w} < \frac{1}{2}(\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)})'\mathbf{w}, \end{aligned} \quad (2.11)$$

con $\mathbf{w} = S^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})$. Il primo termine rappresenta la funzione discriminante lineare basata sui campioni SLDF (sample linear discriminant function) che fu suggerita la prima volta da Fisher nel 1936. La scelta di Fisher fu quella di determinare il

vettore \mathbf{w} come quel vettore tale che l'iperpiano $\mathbf{x}'\mathbf{w}=\text{costante}$ separasse i due campioni in modo da rendere massima la distanza fra le medie dei gruppi e allo stesso tempo minima la varianza all'interno di ciascun gruppo [23]. Egli trovò appunto che tale vettore è proporzionale a $S^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})$.

Esistono vari modi per calcolare la funzione discriminante lineare di Fisher, noi abbiamo usato il metodo di regressione lineare multipla di una variabile *dummy* che descriviamo nel seguito.

Una variabile *dummy* può essere definita come una variabile qualitativa che diventa una finta variabile quantitativa, con la trasformazione dell'informazione circa l'appartenenza al gruppo di controllo oppure al gruppo dei malati in numeri, mediante la loro codifica, ad esempio, in 1 e -1.

In particolare scegliendo la variabile *dummy* seguente:

$$y_{\alpha}^{(1)} = n_2/(n_1 + n_2) \quad (2.12)$$

con $\alpha = 1, \dots, n_1$,

$$y_{\alpha}^{(2)} = -n_1/(n_1 + n_2)$$

con $\alpha = 1, \dots, n_2$, si ottiene una variabile a media nulla sull'intero set di dati.

Si trova la regressione lineare della variabile $y_{\alpha}^{(i)}$ sulle variabili $\mathbf{x}_{\alpha}^{(i)}$ con il metodo OLS determinando il vettore \mathbf{w} che minimizza

$$\sum_{i=1}^2 \sum_{\alpha=1}^{n_i} [y_{\alpha}^{(i)} - \mathbf{w}'(\mathbf{x}_{\alpha}^{(i)} - \bar{\mathbf{x}})]^2, \quad (2.13)$$

dove

$$\bar{\mathbf{x}} = (n_1\bar{\mathbf{x}}^{(1)} + n_2\bar{\mathbf{x}}^{(2)})/(n_1 + n_2). \quad (2.14)$$

L'espressione algebrica del vettore \mathbf{w} che minimizza la forma quadratica in (2.13) si ricava uguagliando a zero le sue derivate parziali rispetto ad ogni componente di \mathbf{w} .

Risultano le seguenti equazioni normali:

$$\begin{aligned} \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})(\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})' \mathbf{w} &= \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} [y_\alpha^{(i)} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})] \\ &\Rightarrow \\ \mathbf{w} &= A^{-1} \mathbf{b} \end{aligned} \quad (2.15)$$

dove $A = \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})(\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})'$ e $\mathbf{b} = \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} [y_\alpha^{(i)} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})]$.

Si noti che il problema della ricerca del minimo ammette un'unica soluzione \mathbf{w} se A non è singolare, ovvero posto $n_1 + n_2 \geq p + 1$, se

$$X = \begin{pmatrix} \mathbf{x}_1^{(1)} - \bar{\mathbf{x}} \\ \vdots \\ \mathbf{x}_{n_1}^{(1)} - \bar{\mathbf{x}} \\ \mathbf{x}_1^{(2)} - \bar{\mathbf{x}} \\ \vdots \\ \mathbf{x}_{n_2}^{(2)} - \bar{\mathbf{x}} \end{pmatrix}, \quad (2.16)$$

con \mathbf{x}_i indichiamo il vettore riga delle p -osservazioni per l'elemento i -esimo del campione, è a rango pieno di colonna (ossia il rango di X è $p + 1$) cosa che accade solo se le sue colonne sono linearmente indipendenti.

Dimostriamo nel seguito che il vettore \mathbf{w} è in effetti proporzionale a $S^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})$.

Partendo ancora una volta dalle equazioni normali risulta infatti:

$$\begin{aligned} \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})(\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})' \mathbf{w} &= \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} [y_\alpha^{(i)} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})] = \\ &= \frac{n_1 n_2}{n_1 + n_2} [(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}) - (\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}})] = \\ &= \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}). \end{aligned} \quad (2.17)$$

Riscrivendo opportunamente la matrice che moltiplica \mathbf{w} :

$$\begin{aligned} \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})(\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})' &= \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}}^{(i)})(\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}}^{(i)})' + \\ &\quad + n_1(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}})(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}})' + n_2(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}})(\bar{\mathbf{x}}^{(2)} - \bar{\mathbf{x}})' = \\ &= \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} (\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}}^{(i)})(\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}}^{(i)})' + \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})', \end{aligned} \quad (2.18)$$

la 2.17 diventa:

$$(n_1 + n_2 - 2)S\mathbf{w} = (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}) \left[\frac{n_1 n_2}{n_1 + n_2} - \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{w} \right]. \quad (2.19)$$

Essendo $(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{w}$ uno scalare, si vede che \mathbf{w} è proporzionale a $S^{-1}(\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})$.

Dalla (2.9), un individuo “futuro” viene collocato nel primo gruppo se il valore previsto per $y_\alpha^{(i)}$, cioè $\mathbf{w}'(\mathbf{x}_\alpha^{(i)} - \bar{\mathbf{x}})$, risulta maggiore o uguale a $\frac{1}{2}(\bar{\mathbf{x}}^{(1)} + \bar{\mathbf{x}}^{(2)} - 2\bar{\mathbf{x}})' \mathbf{w}$, altrimenti nel secondo.

Nel seguito per alleggerire la notazione laddove non è strettamente necessario eviteremo l'indice di appartenenza al gruppo. Indicheremo con \mathbf{x}_i e y_i con $i = 1, \dots, n$ e $n = n_1 + n_2$ rispettivamente il vettore p -dimensionale delle osservazioni e la variabile dummy relative all'elemento i -esimo del campione così ordinato:

$$X = \begin{pmatrix} \mathbf{x}_1^{(1)} - \bar{\mathbf{x}} \\ \vdots \\ \mathbf{x}_{n_1}^{(1)} - \bar{\mathbf{x}} \\ \mathbf{x}_1^{(2)} - \bar{\mathbf{x}} \\ \vdots \\ \mathbf{x}_{n_2}^{(2)} - \bar{\mathbf{x}} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1^{(1)} \\ \vdots \\ y_{n_1}^{(1)} \\ y_1^{(2)} \\ \vdots \\ y_{n_2}^{(2)} \end{pmatrix}. \quad (2.20)$$

3.3 Tecniche di selezione

Il nostro obiettivo è identificare il sottoinsieme delle q caratteristiche del nostro campione p -variato più importanti nel discriminare tra sani e malati. L'approccio

più generale per la selezione di variabili consiste nel definire la funzione obiettivo con la proprietà che il suo massimo o minimo sui sottoinsiemi (di una certa dimensione) delle variabili determini il migliore sottoinsieme e cercare poi tra i vari sottoinsiemi proprio quello che ottimizza questa funzione.

Chiaramente il migliore sottoinsieme di dimensione q può essere trovato esaminando sistematicamente tutti i possibili sottoinsiemi di dimensione q , cioè $\frac{p!}{q!(p-q)!}$ sottoinsiemi, ma tale procedura può essere praticamente irrealizzabile se p è molto grande. Di conseguenza sono state studiate delle strategie per restringere il numero di sottoinsiemi da confrontare. Le strategie più usate sono la *forward selection* (FS) e la *backward elimination* (BE). Entrambi questi metodi costruiscono il sottoinsieme migliore a passi successivi ogni volta eliminando o aggiungendo una variabile.

Nella FS le features sono esaminate a turno per vedere qual'è quella che ottimizza il criterio scelto. A questo punto si considerano tutte le coppie di variabili formate da questa variabile e da ciascuna delle $(p-1)$ variabili rimaste per trovare la coppia che ottimizza il criterio scelto. Si formano a questo punto delle terne con la coppia "migliore" e ciascuna delle $(p-2)$ features rimaste per cercare tra queste la terna "migliore". Si può andare avanti così fino ad ottenere il sottoinsieme q -dimensionale migliore con $q \leq p$ features dopo aver valutato la funzione obiettivo per $p + (p-1) + \dots + (p-q+1) = 1/2 * q * (2p - q + 1)$ volte piuttosto che $\frac{p!}{q!(p-q)!}$ volte come richiesto considerando tutti i sottoinsiemi.

La BE funziona in direzione opposta. Dal set completo delle variabili ogni variabile è eliminata a turno e ogni volta viene valutata la funzione obiettivo per scegliere poi tra i sottoinsiemi di $(p-1)$ variabili quello che ottimizza il criterio scelto. Il processo viene ripetuto e si individua l'insieme q -dimensionale delle feature più importanti anche questa volta dopo $1/2 * q * (2p - q + 1)$ valutazioni della funzione obiettivo [22], [23]. Una delle metodologie più recenti, che procede ugualmente per eliminazioni successive

ma valutando un numero di sottoinsiemi ridotto rispetto alla tecnica BE suddetta, è l'algoritmo Recursive Feature Elimination (RFE), proposto da Guyon et al. nel 2002 [6]. L'algoritmo originale, sviluppato nel caso di classificazione dicotomica, usa la tecnica SVM (support vector machine) con Kernel lineare per identificare, con un processo ricorsivo, le caratteristiche più importanti del campione multivariato. Noi abbiamo utilizzato la procedura indicata da Guyon adattandola al nostro classificatore. Si inizia considerando tutte le variabili. Ad ogni passo si considerano le seguenti operazioni:

- si fitta il classificatore lineare ottenendo i valori dei coefficienti w_i ,
- si ordinano le variabili in gioco in base al valore $\|w_i\|$
- si rimuove la variabile con $\|w_i\|$ minore.

Con questo metodo, per determinare il sottoinsieme q -dimensionale, si calcola solo $(p - q)$ volte la funzione discriminante, risultano quindi drasticamente ridotti i tempi di calcolo.

Osserviamo che la procedura FS è tale che se una feature fa parte della coppia più significativa nella classificazione farà parte anche della terna più significativa e in generale di tutti i sottoinsiemi q -dimensionali con $q > 2$ di feature migliori. Al contrario nelle procedure BE e RFE se una feature non entra a far parte dell'insieme q -dimensionale ($q > 2$) delle feature migliori non farà parte della coppia migliore. Queste caratteristiche potrebbero creare degli effetti non desiderati nella selezione. Ad esempio, supponiamo che le variabili x_i e x_j con $i, j \in \{1, \dots, p\}$ non siano importanti separatamente ma a causa della struttura di correlazione la coppia (x_i, x_j) sia importante. La procedura FS non porterebbe all'inclusione delle due variabili suddette perchè non importanti singolarmente. D'altro canto se tra le coppie (x_i, x_j) ,

(x_i, x_k) e (x_j, x_k) quest'ultima è la coppia più importante il metodo BE esclude la x_i che considerando le tre variabili singolarmente poteva essere, invece, quella più importante. Analogamente nella RFE se usando la terna (x_i, x_j, x_k) risulta w_i il coefficiente minore la variabile x_i che considerata singolarmente poteva essere la più discriminante sarà esclusa già a questo step.

3.4 Il criterio di selezione

Per definire operativamente le varie tecniche di selezione esposte, è necessario scegliere la funzione obiettivo, cioè una quantità con la proprietà che il suo massimo o minimo sui sottoinsiemi (di una certa dimensione) delle variabili determini il migliore sottoinsieme. Una delle funzioni obiettivo che si può definire per individuare il set di variabili più discriminanti è una misura dei tassi d'errore dei diversi classificatori dello stesso tipo implementati sui diversi sottoinsiemi di variabili da confrontare. Tra tutti i sottoinsiemi di variabili di uguale dimensione si può scegliere, ad ogni passo delle procedure su indicate, come set di variabili più discriminanti quello a cui corrisponde il minimo valore dell'errore. Esponiamo nel seguito i metodi più diffusi per valutare l'errore dei diversi classificatori e in particolare il criterio da noi usato.

In generale verificare la validità di un classificatore significa stimare la probabilità di attribuire, mediante la regola di decisione trovata, un'etichetta sbagliata ad un individuo non etichettato.

Il metodo più semplice fornisce una stima di questa probabilità basata sullo stesso set usato per l'addestramento della macchina. Tale metodo consiste nel riclassificare le osservazioni dei due campioni estratti da ciascuna delle due popolazioni tramite la regola di decisione prescelta. Si valuta il *tasso d'errore apparente* calcolando la

frazione di osservazioni classificate erroneamente

$$e = (e_1 + e_2)/n, \quad (4.21)$$

dove e_1 e e_2 indicano rispettivamente il numero delle osservazioni del training set di dimensione n che il classificatore non identifica correttamente nella classe 1 e nella classe 2. Dato che ogni regola di classificazione stimata è basata su una funzione che separa massimamente i due training set, allora gli individui del training set sono proprio quelli che hanno la minima possibilità di essere misclassificati dalla regola trovata. Il tasso d'errore apparente costituisce quindi una stima distorta (troppo ottimistica) del tasso di errore, in particolare tanto più ottimistica quanto più è ridotto il numero di elementi del training set.

Sono stati sviluppati dei metodi che cercano di valutare la capacità di effettiva generalizzabilità della funzione discriminante al di fuori del campione con cui essa viene costruita. Tra questi ricordiamo una procedura di convalida incrociata, o *k-cross validation*. Si divide il campione in k parti, si usano tutte le possibili combinazioni di $(k-1)$ parti per addestrare la macchina. Ogni volta si classificano gli elementi della cappesima parte del campione, la parte esclusa dalla fase di addestramento, e si calcola il tasso di errore. Si fa, infine, una media dei tassi d'errore così calcolati. Dato che solitamente le dimensioni del campione sono ridotte si rischia, con questo metodo, di perdere troppo in potenza escludendo molti dati dalla fase di addestramento.

Si può ovviare a questo inconveniente escludendo ogni volta un solo elemento dal set di addestramento. Si calcolano le funzioni discriminanti su tutti i possibili sottocampioni di $(n-1)$ individui e si classifica, ogni volta, solamente l'individuo escluso. La stima del tasso di errore è ottenuta dividendo per n il numero di osservazioni assegnate a una popolazione diversa da quella di provenienza. Questa procedura, detta *leave-one-out* (letteralmente "lasciare uno fuori"), comporta comunque un peso computazionale elevato dato che si devono calcolare n funzioni di discriminazione per ogni

sottoinsieme di variabili in esame. (Vedremo più avanti com'è possibile ridurre tali calcoli.)

Consideriamo adesso altre stime della validità del classificatore suggerite dalla nostra particolare scelta della regola di classificazione. Essendo il nostro classificatore basato sulla regressione multipla si possono confrontare le validità dei classificatori con diversi sottoinsiemi di osservabili stimando l'ipotesi di linearità.

Il test F e il test t servono per verificare se al variare della variabile p -dimensionale \mathbf{x} si abbia anche una variazione di tipo lineare del valore della y ; quindi se la relazione tra le due variabili possa essere espressa da una retta. Per questo sono chiamati test per la linearità. Ma essi non danno informazioni sull'intensità della relazione, cioè in merito alla quantità d'informazione sulla variabile y che si ricava dalla conoscenza del valore della \mathbf{x} . Questa informazione può essere ottenuta mediante altri indici come il coefficiente di determinazione R^2 e la varianza dei residui. Il coefficiente di determinazione R^2 è la proporzione di variazione totale della variabile dipendente y che è spiegata dalla variabile indipendente \mathbf{x} . Definiamo R^2 come differenza da 1 del rapporto tra la devianza d'errore e quella totale. Indicando, quindi, con y_i il valore sperimentale associato al valore di \mathbf{x} sull'osservazione i -esima, con \hat{y}_i il corrispondente valore stimato o predetto mediante la retta di regressione, il cosiddetto valore teorico, e con \bar{y} la media di y sul campione, possiamo scrivere:

$$R^2 = 1 - \frac{\text{devianza d'errore}}{\text{devianza totale}} = 1 - \frac{\sum_{i=1}^n [y_i - \hat{y}_i]^2}{\sum_{i=1}^n [y_i - \bar{y}]^2}. \quad (4.22)$$

Il coefficiente R^2 serve per misurare quanto della variabile dipendente y sia predetto dalla variabile indipendente \mathbf{x} ; quindi, per valutare l'utilità dell'equazione di regressione nella previsione dei valori della y . Il valore del coefficiente di determinazione è uguale a 0, quando le due variabili sono completamente indipendenti, è uguale ad 1 quando i punti sperimentali stanno tutti sull'iperpiano.

Nel confronto tra classificatori con lo stesso numero di caratteristiche e le stesse dimensioni dei campioni, un classificatore è tanto migliore quanto più R^2 è vicino ad uno, o equivalentemente, quanto più è piccola la quantità detta anche *rischio empirico* o devianza residua (residual mean square), cioè la quantità che è resa minima nel calcolo del coefficiente angolare \mathbf{w} con il metodo OLS:

$$\sum_{i=1}^n [y_i - \hat{y}_i]^2, \quad (4.23)$$

dove

$$\hat{y}_i = \mathbf{w}'(\mathbf{x}_i - \bar{\mathbf{x}}). \quad (4.24)$$

Osserviamo che questa quantità ha un ruolo fondamentale nella regressione lineare. Essa è al numeratore nel calcolo di R^2 e quindi fornisce, almeno in parte, anche quest'informazione. E', inoltre, al denominatore nel test F sulla linearità.

Tanto minore è il valore del rischio empirico o della varianza residua, tanto maggiore è la significatività della regressione e la capacità predittiva della retta.

Quando scegliamo il sottoinsieme migliore come quello a cui corrisponde il classificatore con minore tasso d'errore apparente di misclassificazione non possiamo dire, ad esempio tra due classificatori con lo stesso numero di errate classificazioni a quale dei due corrispondono le misclassificazioni più "errate". Riteniamo che usare come criterio di selezione il minimo rischio empirico piuttosto che il tasso apparente di misclassificazione minore permetta di "premiare" il classificatore per cui punti (\mathbf{x}_i, y_i) del training set risultano più lontani dall'iperpiano separatore.

Come già osservato per il tasso apparente, anche il rischio empirico ha il difetto di sopravvalutare la capacità del classificatore poichè usa per la validazione i campioni del training set. Queste stime possono avere comunque un valore relativo: ad esempio, nel confronto dei diversi classificatori relativi a diversi sottoinsiemi di variabili con la stessa cardinalità per scegliere tra questi il sottoinsieme migliore.

Abbiamo già detto a proposito dell'errore di leave-one-out di classificazione che esso

valuta invece la capacità del classificatore su individui futuri, individui che non sono stati usati per trovare la regola di classificazione. E' interessante notare che esiste un analogo nella regressione che valuta la capacità predittiva della regressione su individui non usati per il calcolo dei coefficienti: la PRESS statistics (Predicted Residual Error Sum of Squares). Con i coefficienti di regressione, calcolati dopo aver escluso dalla matrice X (definita in (2.16)) il vettore \mathbf{x}_j , si calcola il valore della y predetto da tale classificatore per \mathbf{x}_j , che chiamiamo \hat{y}_{-j} . Ripetendo questa procedura per ogni osservazione \mathbf{x}_j con $j = \{1, \dots, n\}$, si calcola l'errore di leave-one-out di regressione o Press come segue:

$$E_{loo} = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_{-j})^2. \quad (4.25)$$

Notiamo che questo calcolo richiede l'implementazione di n classificatori ognuno con un set di training ottenuto escludendo un diverso valore \mathbf{x}_j dal set completo.

Dimostriamo che è possibile a partire dal classificatore che usa il set completo calcolare per ogni \mathbf{x}_j , mediante le \hat{y}_j , le quantità \hat{y}_{-j} e quindi valutare direttamente l'errore E_{loo} .

Ricordiamo che per i punti dati (\mathbf{x}_j, y_j) con $j = \{1, \dots, n\}$ il nostro classificatore restituisce i valori teorici:

$$\hat{y}_j = \mathbf{w}'(\mathbf{x}_j - \bar{\mathbf{x}})$$

o vettorialmente

$$\hat{\mathbf{y}} = X\mathbf{w} = XA^{-1}\mathbf{b} = XA^{-1}X'\mathbf{y}$$

dove $A = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ e \mathbf{y} il vettore delle osservazioni y_j (2.20).

Ponendo $G = XA^{-1}X'$ riscriviamo brevemente:

$$\hat{\mathbf{y}} = G\mathbf{y} \quad e \quad \hat{y}_i = \sum_{j \neq i} G_{ij}y_j + G_{ii}y_i. \quad (4.26)$$

Consideriamo il classificatore sul set di punti

$$\begin{aligned} (\mathbf{x}_j, y_j) \quad \text{con } j \neq i \\ (\mathbf{x}_i, \hat{y}_{-i}) \end{aligned} \quad (4.27)$$

il valore teorico per l'osservazione i -esima risulta dalla (4.26):

$$\sum_{j \neq i} G_{ij} y_j + G_{ii} \hat{y}_{-i}.$$

Dimostriamo, d'altra parte, che questo valore teorico è proprio uguale a \hat{y}_{-i} .

Indicando con \mathbf{w}^* il vettore che minimizza la quantità:

$$E = \sum_{\substack{k=1:N \\ k \neq i}}^n (y_k - \mathbf{w}'(\mathbf{x}_k - \bar{\mathbf{x}}))^2$$

è immediato vedere che lo stesso vettore \mathbf{w}^* minimizza la quantità

$$E^* = \sum_{\substack{k=1:N \\ k \neq i}}^n (y_k - \mathbf{w}'(\mathbf{x}_k - \bar{\mathbf{x}}))^2 + (\hat{y}_{-i} - \mathbf{w}'(\mathbf{x}_i - \bar{\mathbf{x}}))^2$$

Dalla prima segue che:

$$\hat{y}_{-i} = \mathbf{w}^{*'}(\mathbf{x}_i - \bar{\mathbf{x}})$$

e dalla seconda risulta che per il set (4.27), il valore teorico relativo a $(\mathbf{x}_i - \bar{\mathbf{x}})$ è ancora uguale a $\mathbf{w}^{*'}(\mathbf{x}_i - \bar{\mathbf{x}})$.

Risulta quindi:

$$\hat{y}_{-i} = \sum_{j \neq i} G_{ij} y_j + G_{ii} \hat{y}_{-i}.$$

Da cui segue

$$\hat{y}_{-i} = \sum_{j \neq i} G_{ij} y_j + G_{ii} \hat{y}_{-i} + G_{ii} y_i - G_{ii} y_i \quad (4.28)$$

$$\hat{y}_{-i} = \hat{y}_i + G_{ii}(\hat{y}_{-i} - y_i) \quad (4.29)$$

e ancora, sottraendo ad entrambi i membri y_i , risulta:

$$\hat{y}_{-i} - y_i = \frac{\hat{y}_i - y_i}{1 - G_{ii}}.$$

Abbiamo quindi ridotto il calcolo dell'errore di leave-one-out a

$$E_{loo} = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - G_{ii}} \right)^2 \quad (4.30)$$

dove le quantità \hat{y}_i e G_{ii} si riferiscono al classificatore completo.

3.5 Analisi discriminante lineare regolarizzata

Nel tentativo di applicare le tecniche di selezione di cui abbiamo discusso ci siamo scontrati con alcune difficoltà legate al tentativo di implementare un classificatore nel caso di dimensioni del campione ridotte rispetto alle dimensioni dello spazio delle variabili. In questi casi si parla di problema “poveramente” posto se il numero di parametri da determinare, pari al numero di variabili, è confrontabile con il numero di osservazioni e di problema mal-posto quando esso è addirittura inferiore a quest'ultimo.

In questi casi le stime campionarie dei parametri possono essere altamente instabili con varianze elevate (infinite nel caso $n < p$).

Consideriamo le regole di discriminazione lineare ancora nel caso di popolazioni normali multivariate con $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ ma supponendo che le dimensioni dei campioni siano inferiori al numero p di caratteristiche. Quando $n = n_1 + n_2 < p$ la matrice covarianza campionaria comune S è singolare e questo genera naturalmente dei problemi per la funzione discriminante lineare campionaria.

Considerando invece la soluzione ai minimi quadrati ci troveremo di fronte ad una matrice $A = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ singolare ed esisterebbero infiniti iperpiani tali da

soddisfare la condizione di minimo (problema mal posto). In generale un problema è mal posto nel senso di Hadamard se non sono verificate le tre condizioni seguenti: esistenza, unicità e stabilità della soluzione. L'approccio dei minimi quadrati regolarizzati permette di ottenere un'unica soluzione aggiungendo una condizione aggiuntiva alla richiesta di minimizzare la quantità in (2.15). Notiamo che questo metodo permette di risolvere diversi problemi mal posti. Se, ad esempio, le colonne di X sono quasi linearmente dipendenti (problema della collinearità) il sistema risulta ancora mal posto e le stime dei minimi quadrati possono subire ampie modifiche anche per piccole variazioni del campione. Impiegando tecniche di regolarizzazione si ottiene una soluzione con un profilo regolare, liscio (risultati più stabili).

Si ottiene una soluzione regolarizzata secondo il metodo di Tikhonov minimizzando, piuttosto che la (2.15), il funzionale seguente:

$$L(\mathbf{w}) = \sum_{i=1}^n [y_i - \mathbf{w}'(\mathbf{x}_i - \bar{\mathbf{x}})]^2 + \lambda \|\mathbf{w}\|^2, \quad (5.31)$$

dove $\|\mathbf{w}\|$ è la norma euclidea indotta dal prodotto scalare. Il parametro di regolarizzazione λ determina il peso tra la parte relativa al modello funzionale e la parte di regolarizzazione.

Con questo metodo tentiamo di migliorare le stime dei parametri introducendo un criterio che riteniamo fisicamente plausibile che ci condurrà alla scelta di λ .

Mediante tale regolarizzazione si penalizzano i valori grandi di $\|\mathbf{w}\|$. Si può dimostrare, infatti, che se esiste un \mathbf{w}^* che minimizza $L(\mathbf{w})$ (5.31) per un fissato $\lambda > 0$ allora \mathbf{w}^* minimizza:

$$\sum_{i=1}^n [y_i - \mathbf{w}'(\mathbf{x}_i - \bar{\mathbf{x}})]^2$$

con $\|\mathbf{w}\|^2 \leq \|\mathbf{w}^*\|^2$ [24].

Vediamo un'interpretazione geometrica del termine di regolarizzazione.

Consideriamo le coppie ordinate (\mathbf{x}_i, y_i) come punti di uno spazio $(n+1)$ -dimensionale.

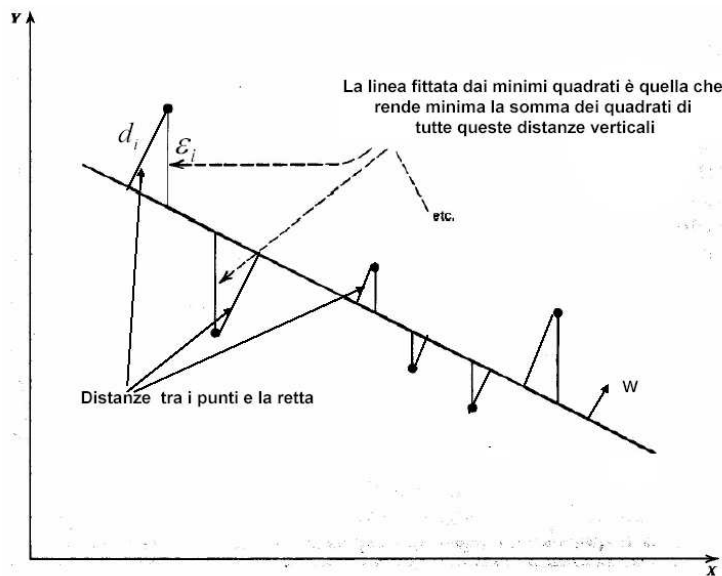


Figura 5.2: Interpretazione geometrica della regolarizzazione

Ogni funzione $y = \mathbf{w}'\mathbf{x}$ rappresenta un iperpiano in questo spazio che approssima i punti dati. L'errore di predizione quadratico sul punto i è così definito:

$$\epsilon_i = (y_i - \mathbf{w}'\mathbf{x}_i)^2.$$

Indichiamo con d_i , invece, la distanza al quadrato tra il punto e l'iperpiano approssimante. Da semplici considerazioni geometriche (figura 3.5) segue:

$$d_i = \frac{\epsilon_i}{1 + \|\mathbf{w}\|^2}.$$

E' quindi evidente che più è piccola la norma di \mathbf{w} , meglio la deviazione ϵ_i approssima la distanza d_i . Concludiamo che il ruolo del termine di regolarizzazione la cui rilevanza dipende dal parametro λ , è permettere di scegliere lo stimatore \mathbf{w} in modo da ottenere l'iperpiano che minimizza la distanza tra se e i punti dati [25].

Analogamente a quanto fatto per i minimi quadrati ordinari, minimizzando il funzio-

nale L si trova l'iperpiano approssimante:

$$\mathbf{w} = (A + \lambda I)^{-1} \mathbf{b} \quad (5.32)$$

dove A e \mathbf{b} sono gli stessi della (2.15).

Ricordiamo che è necessario scegliere un criterio per selezionare il parametro λ .

Dato che il nostro obiettivo è la selezione e la funzione obiettivo scelta è l'errore di leave-one-out, abbiamo scelto il parametro λ come il valore che minimizza l'errore di leave-one-out di regressione (4.30):

$$E_{loo} = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - G_{ii}} \right)^2.$$

E' facile rendersi conto, infatti, che dato che la matrice G , nel caso della funzione discriminante lineare regolarizzata, diventa:

$$G = X(A + \lambda I)^{-1} X',$$

l'errore di leave-one-out attraverso G dipende dal parametro λ .

Capitolo 4

L'analisi dei risultati

Il nostro obiettivo è stato realizzare un classificatore binario “ottimale” per stimare l'appartenenza di un individuo alla classe dei sani o dei malati di CHF.

Nel tentativo di implementare il classificatore è emersa subito l'esigenza di ridurre il numero delle feature. Oltre a problemi legati alla difficoltà dell'applicabilità clinica di un classificatore con molte variabili, un aspetto sostanziale è che le prestazioni del classificatore dipendono dalle dimensioni del campione e dal numero di variabili in maniera congiunta. Implementare un classificatore comporta la stima di alcuni parametri a partire dai valori delle caratteristiche osservate sui campioni di individui delle due classi. Proprio la necessità di stime campionarie, unita a quella di testare alcune ipotesi sui dati, come ad esempio l'omoschedasticità e la multinormalità, lega il numero delle variabili alle dimensioni del campione nel determinare le prestazioni del classificatore. Si dimostra infatti che, a parità di tasso di errore, un classificatore necessita di una sequenza di training che cresce esponenzialmente con il numero delle variabili [1].

Nel nostro caso, avendo a disposizione un campione di $n=303$ osservazioni di $p=400$

variabili, occorre individuare un sottoinsieme di cardinalità $q < p$ dell'insieme delle variabili disponibili che massimizzi le prestazioni del classificatore, cioè che riduca la probabilità di classificare erroneamente un individuo. Abbiamo quindi usato come criterio di selezione diverse stime del tasso d'errore del classificatore.

Il nostro classificatore si basa sulla funzione discriminante di Fisher: cerchiamo l'equazione dell'ipersuperficie che separa i nostri punti in due gruppi in modo che siano massima la differenza tra le medie nei due gruppi e minime le varianze all'interno di ciascun gruppo. Nonostante l'evidenza teorica della necessità di usare una funzione discriminante quadratica nel caso di distribuzioni multivariate normali ma con matrici di dispersione diverse (eteroschedasticità), alcuni studi hanno dimostrato che nel caso di campioni di dimensioni ridotte la funzione discriminante lineare mostra prestazioni superiori [26].

Abbiamo scelto quindi di determinare il classificatore binario basato sulla funzione discriminante lineare di Fisher e, con il metodo dei minimi quadrati OLS (2.15) della regressione della variabile dummy (capitolo 3), abbiamo determinato l'iperpiano separatore.

Abbiamo implementato il classificatore in linguaggio Matlab e applicato diverse strategie di selezione di cui esporremo nel seguito i risultati.

4.1 Le capacità discriminanti individuali delle variabili

Sebbene il nostro obiettivo sia quello di usare simultaneamente le misurazioni delle diverse caratteristiche, ci è sembrato opportuno partire da un'analisi delle capacità discriminanti delle singole variabili.

Ricordiamo che nel nostro caso l'insieme completo delle variabili a disposizione comprende i valori delle entropie campionarie SE e dei coefficienti del modello VAR alle diverse scale temporali e in due diversi regimi di respirazione (basale e controllato) misurati su 256 individui malati di CHF e 47 individui sani.

Le nostre serie sono di circa 600 punti (5 minuti di registrazione con un tasso di campionamento di 2Hz); quindi sono stati scelti fattori di scala $\tau = \{1, \dots, 10\}$ per avere una statistica sufficiente anche alle scale più grandi (60 punti per le serie a $\tau = 10$).

Abbiamo a disposizione quindi in totale un campione di $n=303$ osservazioni di $p=400$ variabili.

Dato che le nostre feature fanno riferimento a grandezze diverse abbiamo ritenuto opportuno scalarle in modo che tutte avessero lo stesso ordine di grandezza così da evitare che feature aventi un ordine di grandezza maggiore rispetto alle altre pesassero di più in fase di classificazione. La tecnica da noi usata a tal fine è stata quella di selezionare la feature i -esima sottrarne la media su tutti i soggetti (pesata sui due insiemi a seconda delle dimensioni del campione, come in (4.24)) e dividerla per la sua deviazione standard (tecnica di standardizzazione).

I valori medi delle feature sui campioni relativi alle due classi con i relativi errori standard sono riportati nelle tabelle in appendice C.

Risultano delle differenze più o meno marcate tra tali valori nelle due classi e, con il test non parametrico di WMW (appendice B), ci siamo preoccupati di valutare la significatività statistica di tali differenze. Nelle tabelle si trovano accanto ai valori medi delle grandezze anche i P-valori del test bilaterale di WMW. Tali valori rappresentano le probabilità che, sotto l'ipotesi H_0 che i valori osservati negli individui delle due classi siano realizzazioni di due variabili casuali con uguali mediane, si osservi il valore della somma dei ranghi calcolato sul campione o uno più estremo. Riteniamo quindi più discriminanti tra le due classi le variabili a cui corrisponde un P-valore

minore e non significative quelle a cui corrisponde un P-valore maggiore di 0.05. Risultano 233 feature significative: sono state scartate il 29% delle entropie e il 45% dei coefficienti del modello VAR.

Abbiamo ordinato tutte le p variabili in ordine decrescente della capacità discriminante, cioè in ordine crescente del P-valore.

Abbiamo valutato l'area ROC relativa ad ogni variabile ed usato un test z per valutare la significatività della capacità discriminante: un valore di $z > 1.96$ indica una prestazione del test significativamente superiore a quella di un test non informativo con $P < 0.05$ (appendice A). Ordinando ancora una volta le feature in ordine decrescente della capacità discriminante, otteniamo lo stesso ordinamento e lo stesso numero di feature significative (a parte sei permutazioni tra due feature consecutive).

La variabile più discriminante è il coefficiente di autocorrelazione lineare della serie DAP a $\tau = 1$ in regime di respirazione basale: il valore assunto da tale variabile è significativamente più alto per i sani (figura 1.1). Per i sani il valore del segnale all'istante t si discosta, in media, meno da quello all'istante $t-1$ indicando una maggiore predicibilità lineare di tali serie nei sani rispetto ai malati. Le tre variabili successive sono ancora variabili relative a serie DAP, nell'ordine, quella in regime basale a $\tau = 2$ e quelle in regime controllato a $\tau = 1$ e $\tau = 2$. Anche per queste il valore è significativamente più alto nei sani.

4.2 Strategie di selezione

L'approccio generale per la selezione delle variabili consiste nel definire una funzione obiettivo con la proprietà che il suo massimo o minimo su tutti i possibili sottoinsiemi di caratteristiche di uguale dimensione q indichi il sottoinsieme migliore q -dimensionale. Per scegliere il sottoinsieme "globalmente" ottimo è necessario poi

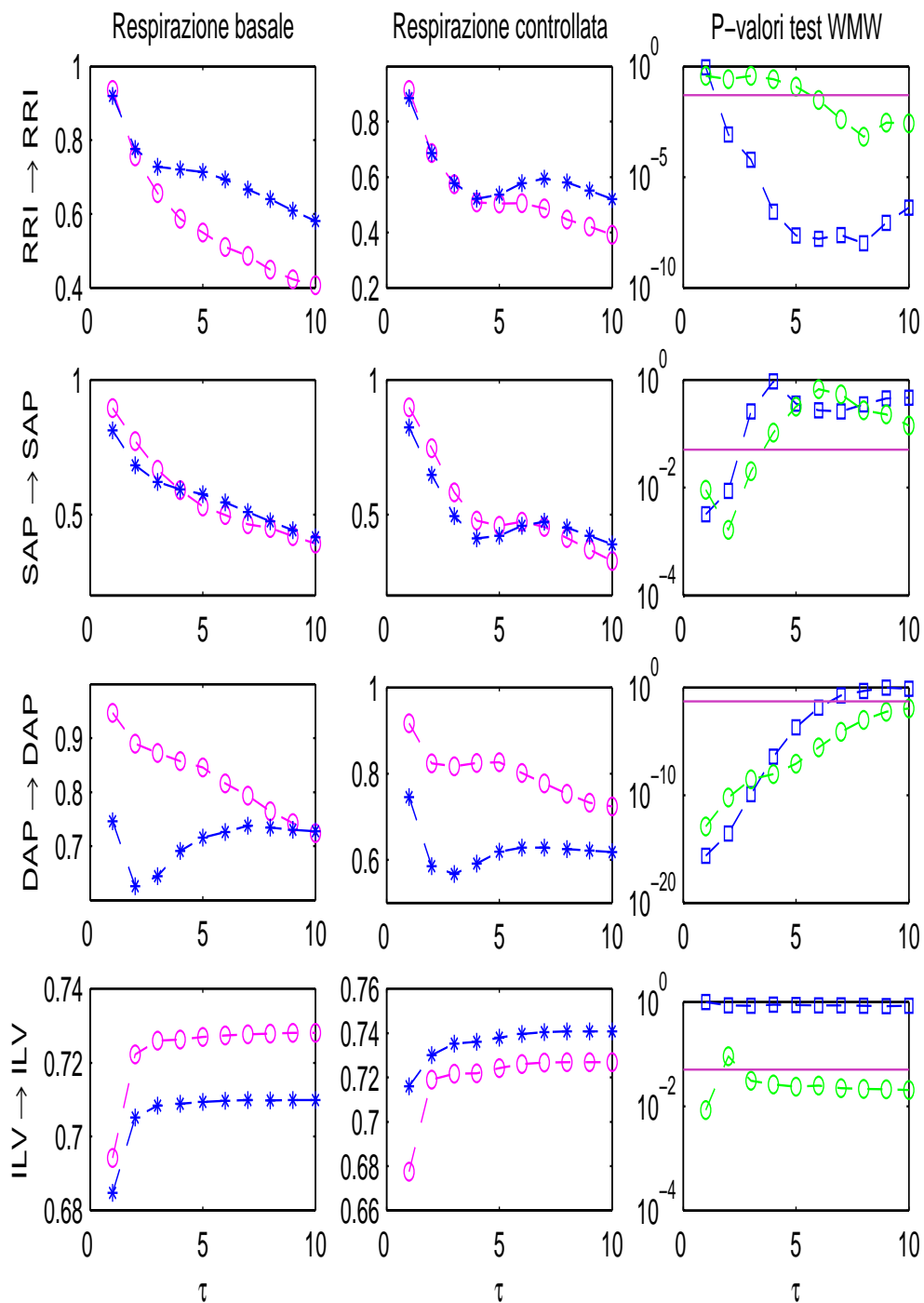


Figura 1.1: Valori medi dei coefficienti di autocorrelazione lineare in regime di respirazione basale (prima colonna) e controllata (seconda colonna) sul gruppo dei sani (sferette rosa) e sul gruppo dei malati (stelle blu), P-valori del test WMW tra i due gruppi in regime basale (quadretti blu) e in regime controllato (sferette verdi) e valore di soglia di $P=0.05$ (linea viola) (terza colonna) alle diverse scale temporali τ .

confrontare tra loro tutti i sottoinsiemi con $q = \{1, \dots, p\}$.

In un primo studio abbiamo usato come funzione obiettivo delle diverse strategie di selezione due semplici stime dell'errore del classificatore basate sui dati di training: il tasso d'errore apparente di classificazione (4.21) e il rischio empirico (4.23).

Come già osservato nel capitolo 3 una ricerca esaustiva del miglior sottoinsieme q -dimensionale tra tutti i possibili sottoinsiemi comporterebbe il confronto di $\frac{p!}{q!(p-q)!}$ sottoinsiemi e, nel nostro caso, avendo a che fare con $p=400$ variabili il problema sarebbe intrattabile. Abbiamo sviluppato diverse strategie che riducono il numero di confronti da effettuare ma che assumono approssimazioni più o meno marcate nelle diverse regioni di variabilità di q .

La prima strategia di selezione adottata è basata sulla capacità discriminante delle variabili prese singolarmente. Sono stati implementati p classificatori utilizzando ogni volta una delle variabili disponibili. Abbiamo ottenuto due ordinamenti delle feature sistemandole in ordine decrescente della capacità di classificazione valutata attraverso le due stime d'errore suddette.

Notiamo per inciso che ordinando le feature secondo l'ordine decrescente della distanza di Mahalanobis si ottiene lo stesso ordinamento ottenuto dalla stima dell'errore rischio empirico [23]. Tale distanza è definita come

$$\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (2.1)$$

ed è un indice della distanza tra le due popolazioni.

Per ogni classificatore abbiamo stimato tale distanza mediante le semplici medie campionarie e la matrice di covarianza campionaria comune.

In entrambi gli ordinamenti trovati sono state raggruppate le caratteristiche in sottoinsiemi q -dimensionali prendendo le prime q feature. Per ciascun sottoinsieme è stato implementato un classificatore e sono stati calcolati: il tasso d'errore apparente di classificazione e il rischio empirico. Ci siamo limitati a un valore superiore di 302

per q data l'impossibilità di determinare con il metodo OLS un unico iperpiano, e quindi un classificatore lineare, quando $n \leq q$.

Nei grafici nelle figure (2.2) e (2.3) sono riportati rispettivamente il tasso d'errore apparente e il rischio empirico in funzione delle dimensioni del relativo sottoinsieme (curve verdi nelle figure).

Questa strategia assume l'approssimazione che le variabili più discriminanti singolarmente lo siano anche in associazione con le altre, approssimazione che è tanto più grossolana quanto maggiore è la dimensione del sottoinsieme in esame.

Ricordiamo che per la determinazione del migliore sottoinsieme q -dimensionale esistono delle procedure euristiche passo-passo molto diffuse: la FS e la BE che sono rispettivamente una procedura di selezione in avanti e una di eliminazione all'indietro e che fanno assunzioni meno forti rispetto al raggruppamento secondo le capacità individuali appena visto. Entrambe riducono la selezione al confronto di $\frac{1}{2}q(2p - q + 1)$ sottoinsiemi ma hanno, nel campo di variabilità di q , regioni di approssimazione opposte. La procedura FS, che procede in ordine di q crescente ipotizzando ad ogni passo che un indicatore nell'insieme q -dimensionale migliore appartenga all'insieme $(q+1)$ -dimensionale migliore, dà soluzioni più approssimate per valori grandi di q ; al contrario la procedura di eliminazione all'indietro è più approssimata per piccoli valori di q . Come già detto nel capitolo 3 esiste una procedura più recente, la tecnica RFE, che riduce ulteriormente il numero di confronti. E' richiesta, infatti, l'implementazione di soli $(p-q)$ classificatori per determinare il migliore sottoinsieme q -dimensionale.

Le procedure di eliminazione all'indietro partono dall'implementazione di un classificatore p -dimensionale ma ciò richiederebbe nel nostro caso, dato che $p > n$, un campione più numeroso. Abbiamo quindi usato la tecnica FS individuando i migliori

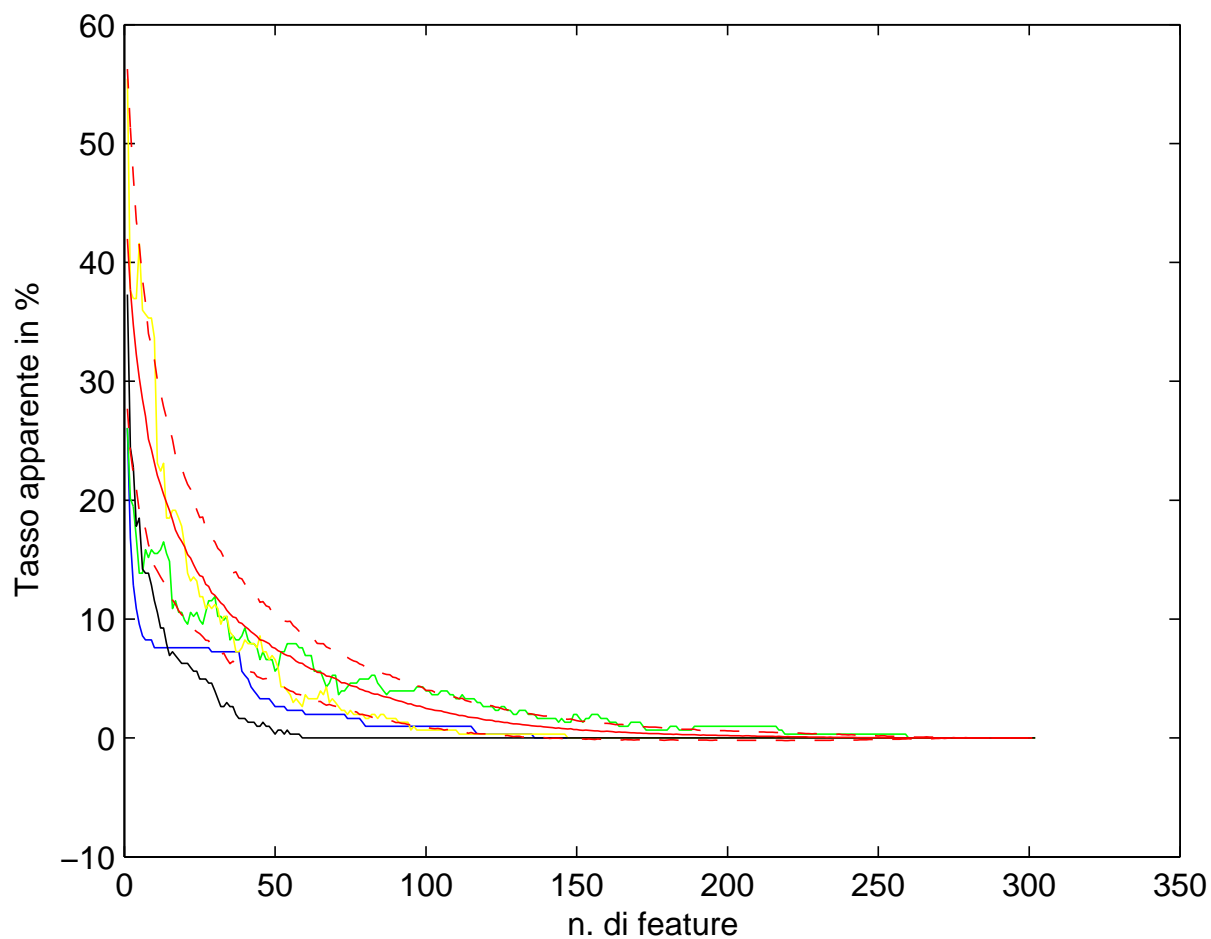


Figura 2.2: Tassi di errore apparente di classificazione al variare della dimensione n dei sottoinsiemi individuati mediante le diverse procedure di selezione: metodo univariato (curva verde), FS (curva blu), RFE (curva gialla), BE (curva nera), e fascia random con intervallo di confidenza del 95% (curve rosse)

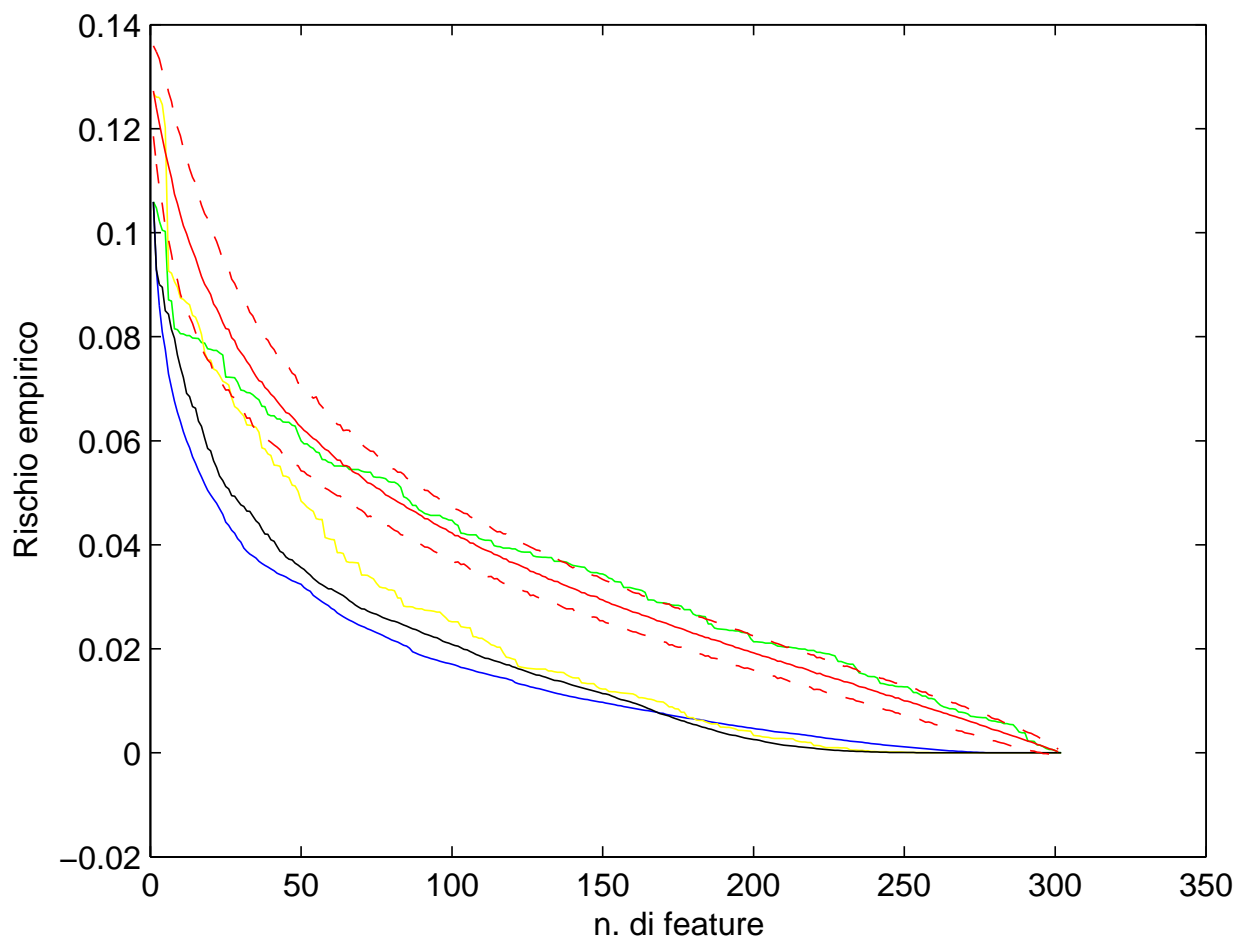


Figura 2.3: Rischio empirico al variare della dimensione n dei sottoinsiemi individuati mediante le diverse procedure di selezione: metodo univariato (curva verde), FS (curva blu), RFE (curva gialla), BE (curva nera), e fascia random con intervallo di confidenza del 95% (curve rosse).

sottoinsiemi q -dimensionali con $q = \{1, \dots, 302\}$, cioè $q < n$, usando ancora una volta le due stime d'errore suddette.

Osserviamo che esiste una funzione di Matlab che fornisce un'interfaccia grafica interattiva che permette di realizzare una procedura di "stepwise regression". In generale una procedura di stepwise al primo step sceglie la prima variabile da includere nel modello e successivamente ad ogni step procede per inclusione, esaminando però, di volta in volta, l'insieme delle variabili ottenute per individuarne, eventualmente, alcune che possono essere omesse senza sostanziale peggioramento della classificazione. La funzione di Matlab include le variabili con i coefficienti di regressione maggiori purchè siano significativamente diversi da zero con $P\text{-valore} < 0.05$ ed esclude le variabili del set corrente con i coefficienti minori purchè non nulli con $P > 0.10$. Per valutare la significatività viene usato un test-t con gradi di libertà pari a $(n-p)$ in cui il valore di t è definito dal rapporto tra la stima del coefficiente e il suo errore standard. Utilizzando questo metodo risulta un insieme di 30 feature più significative ed è interessante notare che queste si succedono nello stesso ordine ottenuto con il nostro programma per la FS minimizzando il rischio empirico.

Considerando il sottoinsieme di $q=302$ caratteristiche individuato mediante FS abbiamo riordinato secondo le procedure BE e RFE tale sottoinsieme utilizzando le due stime suddette (figure (2.2) e (2.3)) .

Non avendo condotto una ricerca esaustiva su tutti i possibili sottoinsiemi q -dimensionali è legittimo chiedersi quanto i sottoinsiemi ottenuti siano migliori rispetto a sottoinsiemi estratti casualmente dall'insieme totale.

Nelle stesse figure sono riportate le rispettive curve d'errore ottenute mediando su 1000 permutazioni random delle variabili e il corrispondente intervallo di confidenza del 95%. Si osserva che il metodo basato sulle capacità individuali delle caratteristiche dà luogo a curve che si sovrappongono alla fascia delle curve random per cui

risulta del tutto inaffidabile (tranne valori di $q < 10$) per entrambe le stime d'errore. La curva RFE si sovrappone alla fascia random quando si considera il tasso d'errore apparente mentre tale metodo risulta valido quando si basa sulla stima dell'errore rischio empirico almeno per grandi valori di q (per $q > 30$). Ciò conferma quanto detto nel capitolo precedente circa la maggiore affidabilità di quest'ultima misura d'errore. Risulta inoltre evidente che effettivamente i metodi di eliminazione all'indietro individuano dei sottoinsiemi migliori per valori grandi di q e, tra i due, il metodo BE risulta più preciso su tutto il campo di variabilità di q . In particolare nel caso dell'errore apparente di misclassificazione la regione in cui il metodo BE individua il sottoinsieme migliore è molto estesa (da $q=16$ a $q=302$), mentre per q piccoli ($q < 16$) è preferibile il metodo FS. Quando si confronta invece il rischio empirico il metodo FS risulta il più preciso in una zona di variabilità di q più estesa ($q > 160$).

4.3 Metodo di leave-one-out

Nelle procedure descritte precedentemente emergono due tipi di problemi. Il primo deriva dall'aver adottato delle stime dell'errore calcolate sulle sequenze di training e ciò comporta una sottostima dell'errore. Queste stime, che possono avere un valore relativo nel confronto di sottoinsiemi di uguali dimensioni, non permettono di confrontare sottoinsiemi di dimensioni diverse dato che la stima è tanto più ottimistica quanto più cresce il numero di variabili considerate. Cresce con il numero di variabili l'adattamento del classificatore ai nostri dati e le stime degli errori basate sulla sequenza di training perdono sempre più il carattere di generalizzabilità. Esse non sono quindi utili all'individuazione della cardinalità e degli elementi del sottoinsieme multidimensionale più discriminante. Abbiamo allora deciso di considerare come criterio di selezione l'errore di leave-one-out, in particolare l'errore di leave-one-

out di regressione definito nella (4.25), per valutare l'effettiva generalizzabilità della funzione lineare al di fuori del campione con cui essa viene costruita.

Il secondo problema sta nell'impossibilità di implementare il classificatore con $n < q$ con il metodo OLS e nella consapevolezza che anche le soluzioni trovate nel caso in cui n è poco maggiore di q sono molto instabili cioè variano molto al variare dei campioni in esame. Per poter meglio valutare la capacità di un classificatore quando le sue dimensioni diventano confrontabili o addirittura superiori alla dimensione del campione abbiamo usato il metodo dei minimi quadrati regolarizzati (5.32). In particolare abbiamo scelto il parametro di regolarizzazione come quello che minimizza l'errore di leave-one-out.

Come discusso precedentemente il metodo univariato non è affidabile; pertanto l'abbandoniamo. Con le tecniche FS, BE e RFE mediante il nuovo criterio di selezione individuiamo nuove sequenze di sottoinsiemi migliori i cui errori sono riportati nelle curve in figura (3.4). Innanzitutto osserviamo che le curve così ottenute risultano molto più distanti dalla fascia random rispetto alle curve in figura (2.3) ottenute utilizzando il rischio empirico. Ancora una volta è evidente che il metodo BE è il più efficace per grandi valori di q ($q > 110$) mentre il metodo FS lo è per piccoli valori ($q < 50$). Nella regione intermedia il metodo RFE fornisce le soluzioni migliori.

Le curve relative ai metodi FS, BE, RFE presentano ciascuna un minimo rispettivamente in corrispondenza di $q = 168, 137, 96$.

Nel tentativo di estrarre un set di variabili che riassume al meglio le capacità di classificazione dell'intero insieme di partenza, abbiamo, quindi, trovato tre diverse soluzioni che condividono tra loro 39 variabili. Questo sottoinsieme intersezione include il 25% di tutte le entropie e solo il 6% di tutti i coefficienti VAR.

Osserviamo che tra i coefficienti VAR in questo sottoinsieme non ce n'è nessuno di quelli relativi all'influenza delle variazioni di SAP e DAP sulla serie RRI, nè di quelli

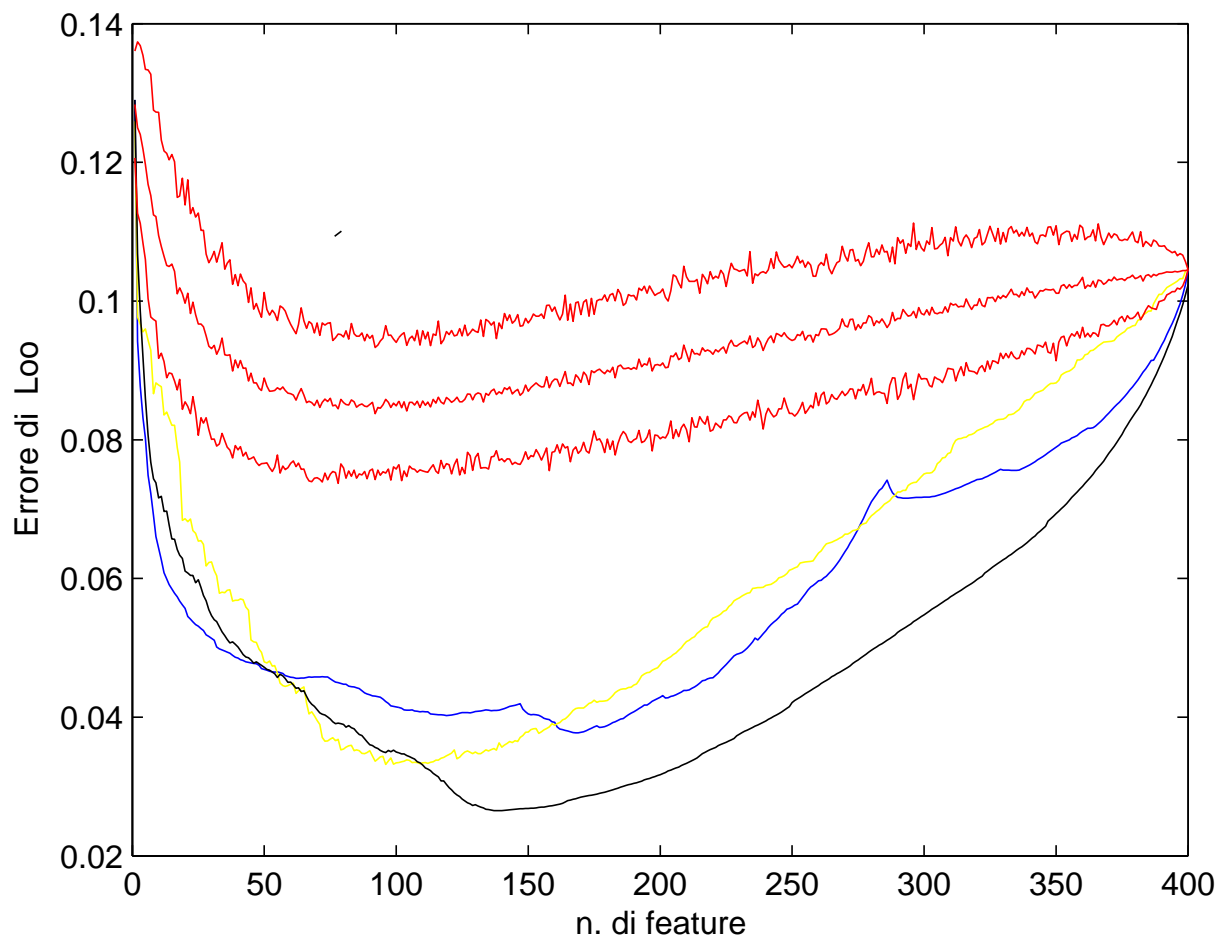


Figura 3.4: Errore di leave-one-out al variare della dimensione n dei sottoinsiemi individuati mediante le diverse procedure di selezione: FS (curva blu), RFE (curva gialla), BE (curva nera), e fascia random con intervallo di confidenza del 95% (curve rosse).

dell'interazione tra SAP e DAP. Il sottoinsieme intersezione include, invece, i coefficienti VAR relativi all'influenza di RRI su DAP in regime di respiro basale a $\tau = 8$ e controllato a $\tau \geq 8$; non ci sono, invece, coefficienti relativi all'influenza di RRI su SAP.

Il sottoinsieme delle feature in comune contiene inoltre i coefficienti dell'influenza di ILV in regime di respirazione basale sulle variazioni di RRI ($\tau = 6, 8$) e di SAP ($\tau = 8, 10$) e alcuni delle altre serie su ILV soprattutto in condizioni di respiro controllato. Sono inclusi anche alcuni coefficienti di autointerazione: quello di RRI ($\tau = 3, 4$), di SAP ($\tau = 6$), DAP ($\tau = 1, 4$) in regime di respirazione basale e di ILV ($\tau = 1$) in regime di respirazione controllata.

E' interessante osservare che il metodo BE e il metodo FS indicano entrambi come miglior classificatore unidimensionale il coefficiente di autocorrelazione lineare della serie DAP a $\tau = 1$ e in regime basale. Questo significa che, in base al nostro studio, tale variabile dovrà appartenere al classificatore migliore qualunque sia la sua dimensionalità.

Per confrontare tra loro i tre classificatori "subottimi" ottenuti abbiamo confrontato le aree ROC delle variabili le cui osservazioni sono i valori di leave-one-out predetti da ciascun classificatore per ogni individuo. Abbiamo studiato la significatività delle differenze a due a due tra le tre aree ROC calcolata rapportando la differenza delle due aree al suo errore standard (vedi Appendice C). Abbiamo calcolato il valore della variabile standardizzata z

$$z = \frac{AUC_A - AUC_B}{\sqrt{\sigma_A^2 + \sigma_B^2}},$$

dove con AUC_A , AUC_B , σ_A^2 , σ_B^2 intendiamo le aree ROC e le rispettive varianze di ciascuna coppia dei tre classificatori migliori. Le aree ROC risultano:

$$AUC_{RFE} = 0.9987 \pm 0.0036 \quad AUC_{BE} = 0.99975 \pm 0.0016 \quad AUC_{FS} = 0.9999 \pm 0.0009$$

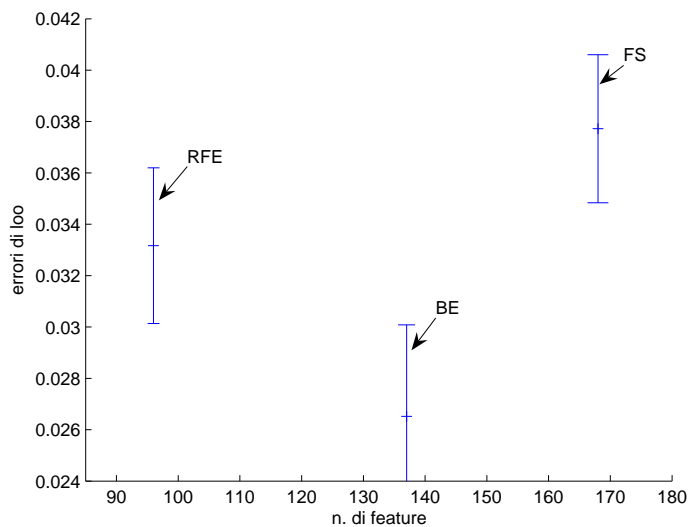


Figura 3.5: I sottoinsiemi migliori

Per tutte e tre le differenze considerate z risulta inferiore a 1.96 e quindi le differenze tra le aree dei sottoinsiemi subottimi non sono significative con $P\text{-valore} < 0.05$.

Abbiamo pensato di adottare il criterio di massima semplicità e scegliere, tra i tre sottoinsiemi individuati, come sottoinsieme migliore quello a cui corrisponde il minimo q purché il suo errore sia compreso nella barra d'errore standard del sottoinsieme con in assoluto il valore più piccolo d'errore (figura 3.5).

Il sottoinsieme trovato con RFE di cardinalità 96 è per noi il sottoinsieme “globalmente” ottimo.

Conclusioni

Nel tentativo di sviluppare un metodo per l'identificazione di pazienti affetti da CHF, è stato costruito un classificatore basato sull'analisi discriminante lineare di Fisher utilizzando la regressione di una variabile dummy sulle caratteristiche in esame. Il problema fondamentale affrontato nella costruzione di tale classificatore riguarda la selezione delle feature più significative. Una selezione permette, infatti, di ridurre i tempi e i costi delle misure, ma soprattutto si rivela necessaria quando il campione disponibile non è sufficientemente numeroso, come nel caso in esame ($n=303$ osservazioni di $p=400$ variabili).

Non è sempre vero, infatti, che un classificatore è tanto migliore quanto maggiore è il numero di variabili considerate perchè noi consideriamo una regola di classificazione che è basata sul campione in esame e che, quando il numero di feature supera un certo valore, comincia a perdere in generalizzabilità quanto più tale numero si avvicina alla cardinalità del campione (overfitting). Questo è evidente confrontando l'errore di leave-one-out per classificatori di diverse dimensioni, piuttosto che l'errore apparente il quale, invece, decresce sempre all'aumentare del numero di variabili.

Per individuare il classificatore ottimo avremmo dovuto confrontare le capacità di classificazione di 2^p sottoinsiemi di variabili. Questo problema con 400 variabili è intrattabile, abbiamo quindi usato diverse procedure di selezione che riducono il numero dei confronti.

Paradossalmente, l'eliminazione di variabili che singolarmente presentano un basso indice di separazione può rivelarsi una pessima idea. Viceversa, è possibile che una variabile con un alto contenuto informativo ai fini della classificazione sia superflua se utilizzata con altre variabili.

Anche nel nostro studio, eliminando via via dal classificatore le variabili singolarmente poco significative sono stati individuati classificatori non molto migliori rispetto a quelli ottenuti con variabili scelte a caso. Le tecniche FS (forward selection), BE (backward elimination) e RFE (Recursive Feature Elimination), che rappresentano un miglior compromesso tra il semplice raggruppamento secondo le capacità individuali discriminanti e la ricerca esaustiva tra tutti i possibili sottoinsiemi, si sono rivelate più efficaci.

Nel tentativo di implementare tali tecniche ci siamo scontrati, tuttavia, con la difficoltà di determinare funzioni discriminanti lineari nel caso in cui il numero delle feature supera quello dei pazienti (problema mal-posto). Ciò ha reso necessario l'uso di un metodo dei minimi quadrati regolarizzati grazie al quale è stato possibile determinare univocamente gli iperpiani separatori anche nel caso di classificatori di dimensioni superiori a quelle del campione e di ottenere soluzioni più stabili per problemi “poveramente” posti (dimensioni del classificatore confrontabili con quelle del campione).

Applicando le tre tecniche di selezione FS, BE e RFE usando come funzione obiettivo l'errore di leave-one-out sono stati individuati altrettanti sottoinsiemi migliori aventi cardinalità (rispettivamente 168, 137, 96) ed errori di leave-one-out diversi. Tali classificatori “migliori” non risultano significativamente diversi mediante il confronto delle aree ROC. Abbiamo, comunque, osservato che in base ad un criterio di massima semplicità il sottoinsieme migliore in assoluto risulta il sottoinsieme individuato da RFE che contiene 96 feature.

Osserviamo che nel tentativo di estrarre un set di variabili che riassume al meglio le capacità di classificazione dell'intero insieme di partenza abbiamo trovato tre diverse soluzioni, tutte ugualmente valide dal punto di vista della classificazione, ma che condividono solo poche variabili fra loro (precisamente 39). Questo inconveniente, detto mancanza di stabilità della soluzione deve essere sempre ben presente in fase di analisi. Sottostimarne l'importanza porta a un falso senso di sicurezza nei confronti del modello individuato dagli algoritmi di selezione.

Ovviamente il problema principale che la mancanza di stabilità pone al ricercatore è che essa causa seri dubbi sull'interpretazione medica dei risultati. Per stimare quanto è generalizzabile e valido il risultato trovato si potrebbe continuare l'analisi ricorrendo a tecniche bootstrap. Queste consistono nel ripetere il processo di selezione su campioni ricavati dal set di dati originario mediante campionamento con reinserimento dei dati, controllando poi quante volte le caratteristiche selezionate sul campione di partenza sono selezionate anche nei campioni bootstrap. Tale procedura non costituirebbe una cura al problema, che deriva dal fatto che le tecniche di selezione sono troppo strettamente legate al campione in esame, ma almeno sarebbe una spia di quanto possa essere ritenuto valido e generalizzabile un risultato.

L'unica vera cura consiste nell'allargare il campione in esame, nella speranza di riuscire a isolare un comportamento generale da quello dovuto a semplice casualità.

Dal confronto con gli errori di leave one out calcolati per i classificatori lineari individuati passo passo mediante le diverse strategie e quelli calcolati su classificatori delle stesse dimensioni ma con feature estratte a caso tra le p variabili in esame e consapevoli del problema della mancanza di stabilità, possiamo concludere che l'insieme da noi identificato è almeno un sottoinsieme soddisfacente seppure non quello globalmente ottimo tra tutti i possibili sottoinsiemi.

Appendice A

Le curve R.O.C.

In medicina clinica vengono spesso utilizzati test che forniscono risultati sotto forma di variabili continue unidimensionali. Supponiamo che le due distribuzioni X_s e X_m dei valori assunti da x rispettivamente in due campioni random delle popolazioni dei sani e dei malati siano quelle rappresentate in figura 0.1. Nel nostro caso il valore di x è tendenzialmente più alto nei malati. Nel caso di variabile continua è necessario individuare un valore di soglia, cut off, (indicato dalla linea rossa in figura) allo scopo di decidere se, in base al valore assunto dalla nostra variabile x , l'individuo debba essere classificato come sano o malato, "negativo" o "positivo". Nella pratica non esiste un cut off che consenta una classificazione perfetta dato che si verifica sempre una sovrapposizione più o meno ampia delle due distribuzioni : ci sarà un certo numero di malati per cui la diagnosi risulterà negativa (falsi negativi) e un numero di sani con diagnosi positiva (falsi positivi).

Il confronto fra i risultati del test in esame e l'autentico stato di ogni individuo consente di stimare la capacità diagnostica del test in esame attraverso due parametri: la sensibilità S_e e la specificità S_p , definite rispettivamente come la probabilità che

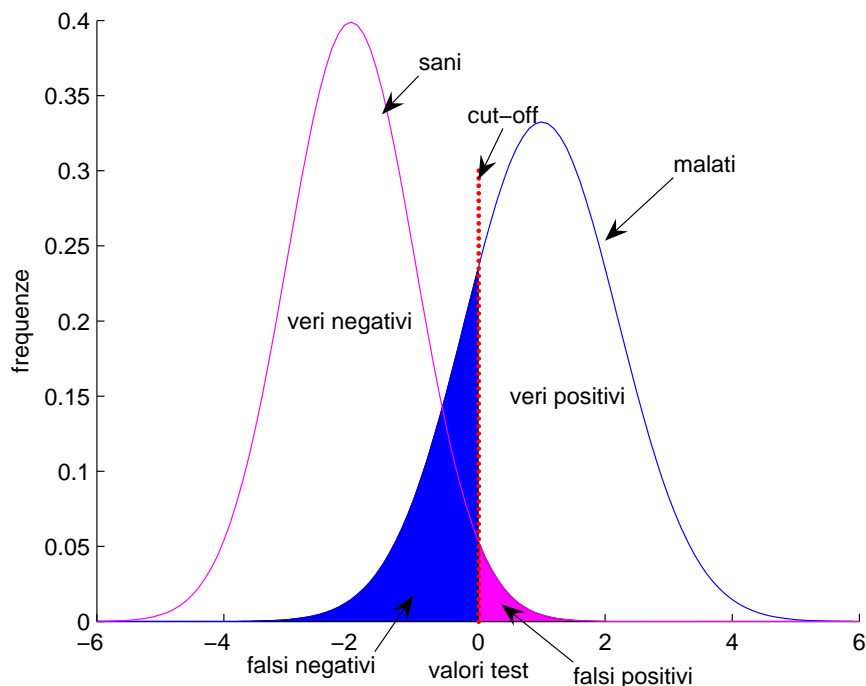


Figura 0.1: Distribuzione degli esiti di un ipotetico test nelle classi di individui sani e malati

un soggetto malato risulti positivo e uno sano negativo cioè:

$$S_e = a/(a + c) \quad (0.1)$$

$$S_p = d/(d + b), \quad (0.2)$$

dove a è il numero di veri positivi, c il numero di falsi negativi, d quello dei veri negativi e b quello dei falsi positivi. Questi parametri dipendono ovviamente dalla scelta del valore di soglia: ogni variazione del valore di cut off adottato coincide con una variazione della sensibilità e della specificità del test in direzioni opposte, nel senso che ad un incremento della sensibilità corrisponde un decremento della specificità e viceversa.

Le curve ROC (Receiver Operating Characteristics curve) offrono un valido aiuto nella scelta del cut off ottimo per il test usato oppure per confrontare tra loro le performance

di due test attraverso l'analisi della correlazione tra sensibilità e specificità per una gamma di differenti valori di cut off.

Tali curve vengono generate riportando in un sistema di assi cartesiani in ascisse la proporzione di falsi positivi ($1 - S_p$) e sulle ordinate quella di veri positivi S_e corrispondenti a diversi valori di cut off. L'unione dei punti ottenuti riportando nel piano cartesiano ciascuna coppia $(S_e, 1 - S_p)$ genera una curva spezzata con andamento a scala interpolando la quale si ottiene la curva Roc.

La capacità discriminante di un test è proporzionale all'area sottesa alla curva (AUC= Area under Curve) ed equivale alla probabilità che la variabile x assuma per un individuo estratto a caso dal gruppo dei malati un valore più alto rispetto ad uno estratto a caso dal gruppo dei sani ¹.

Nel caso di un test perfetto (capacità discriminante 100%), la AUC passa per il punto $(0, 1)$ e la curva delimita l'intero quadrato di area unitaria. Al contrario, la Roc per un test assolutamente privo di valore informativo coincide con la diagonale che passa per l'origine con $AUC = 0.5$. Tutti gli altri casi di accuratezza intermedia sono compresi tra queste due curve.

Il calcolo dell'AUC per una curva empirica (cioè ottenuta da un campione finito) si può ottenere con diversi metodi come la regola trapezoidale o con metodi di stima dell'area "vera" (massima verosimiglianza) e di interpolazione della curva. In pratica si usano dei programmi al computer che calcolano le AUC.

Esiste una rozza classificazione della capacità diagnostica di un test in base ai valori di AUC proposta da Swets (1998):

- $AUC = 0.5$ test non informativo
- $0.5 < AUC \leq 0.7$ test poco accurato

¹ Bamber, 1975

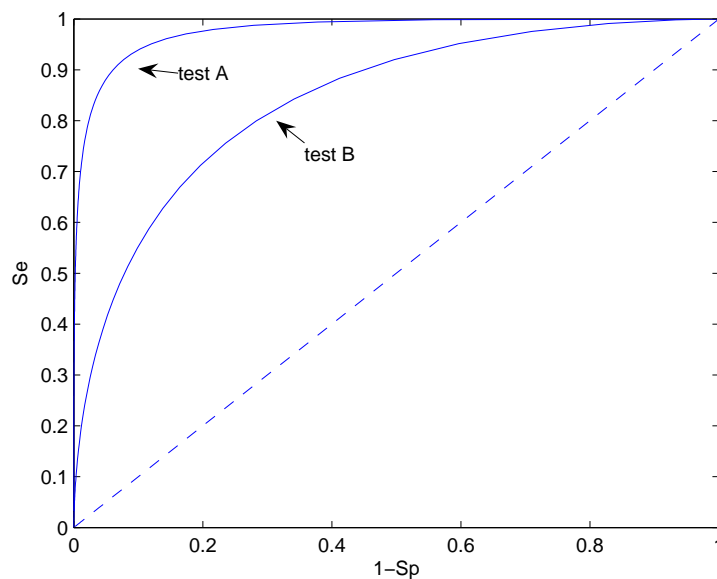


Figura 0.2: Curve Roc relative a due ipotetici test : il test A ha una capacità discriminante maggiore rispetto al test B

- $0.7 < AUC \leq 0.9$ test moderatamente accurato
- $0.9 < AUC \leq 1$ test altamente accurato
- $AUC = 1$ test perfetto.

Altrimenti considerando AUC una variabile normale è possibile costruire un test di significatività statistica z (basato sulla distribuzione normale standardizzata) come segue:

$$z = \frac{AUC - 0.5}{\sqrt{\sigma_{AUC}^2}} \quad (0.3)$$

con σ_{AUC}^2 che rappresenta la varianza di AUC.

Secondo Hanley e McNeil (1983), la varianza dell'area AUC può essere stimata da:

$$\sigma_{AUC}^2 = \frac{AUC(1 - AUC) + (n_1 - 1)(Q_1 - AUC^2) + (n_2 - 1)(Q_2 - AUC^2)}{n_1 n_2} \quad (0.4)$$

dove n_1 e n_2 rappresentano le dimensioni dei due campioni a confronto e $Q_1 = \frac{AUC}{2-AUC}$ e $Q_2 = \frac{2AUC^2}{1+AUC}$ rispettivamente le probabilità che due soggetti presi random dal gruppo dei malati abbiano entrambi valori di x superiori a quelli dei sani e viceversa.

Se il valore di z supera il valore critico di 1.96, si può dire che il test ha una performance significativamente superiore a quella di un test non informativo con $p < 0.05$.

Per confrontare tra loro la validità di due test diagnostici A e B è necessario verificare che la differenza tra le due aree AUC_A e AUC_B sia statisticamente significativa.

Nell'ipotesi bi-normale, un test z può essere eseguito rapportando la differenza delle due aree all'errore standard di questa differenza, cioè:

nel caso di test indipendenti

$$z = \frac{AUC_A - AUC_B}{\sqrt{\sigma_A^2 + \sigma_B^2}}, \quad (0.5)$$

per test non indipendenti (situazione che può verificarsi quando essi vengono applicati agli stessi soggetti)

$$z = \frac{AUC_A - AUC_B}{\sqrt{\sigma_A^2 + \sigma_B^2 - 2r\sigma_A\sigma_B}}, \quad (0.6)$$

con r parametro di correlazione tra le due AUC ricavato a partire dai coefficienti di correlazione dei valori dei due test nei due gruppi mediante le tavole di Hanley e McNeil [27].

Appendice B

Il test di Wilcoxon-Mann-Whitney della somma dei ranghi

Il test di Wilcoxon-Mann-Whitney della somma dei ranghi (WMW test) è uno dei test non parametrici per due campioni indipendenti utilizzati per inferenze sulla tendenza centrale. L'applicazione del metodo richiede che

- le due popolazioni a confronto siano distribuite in modo continuo,
- siano entrambe simmetriche o entrambe asimmetriche nello stesso modo,
- i dati siano misurati con una scala almeno ordinale.

Supponiamo di voler confrontare il gruppo A e il gruppo B rispetto ad una variabile avendone raccolto rispettivamente n_A misure per il primo e n_B per il secondo.

Essendo Me_A e Me_B le rispettive mediane, le ipotesi che il test WMW può verificare possono essere:

- **bilaterali** $H_0 : Me_A = Me_B$ contro $H_1 : Me_A \neq Me_B$,
- **unilaterali**
 - **sinistra** $H_0 : Me_A \geq Me_B$ contro $H_1 : Me_A < Me_B$
 - oppure **destra** $H_0 : Me_A \leq Me_B$ contro $H_1 : Me_A > Me_B$

La procedura è la seguente:

- si combinano i dati dei due gruppi in una serie unica disponendo i valori in ordine crescente e conservandone l'informazione relativa al gruppo di appartenenza,
- si assegna il rango ad ogni elemento, cioè un punteggio da 1 in su a partire dal valore più basso, (nel caso di valori uguali si assegna ad ognuno un punteggio pari alla media dei ranghi che altrimenti sarebbero stati assegnati),
- si calcola la somma dei ranghi (chiamata w) del gruppo con il numero di dati minore,
- infine la significatività della differenza tra le mediane dei due gruppi può essere valutata confrontando il valore w calcolato con la distribuzione di probabilità della corrispondente variabile aleatoria W nell'ipotesi H_0 [28].

Quando l'ipotesi nulla H_0 del test bilaterale è vera, i valori del gruppo A sono casualmente mescolati con quelli del gruppo B, di conseguenza, per quanto concerne i ranghi (e quindi W) è come se noi etichettassimo con A n_A valori scelti casualmente tra gli $n = n_A + n_B$ a disposizione e con B i rimanenti. Esistono programmi informatici che calcolano la distribuzione di frequenza di tutti i possibili valori W , ottenibili per ogni combinazione di ranghi nel gruppo di n osservazioni.

Se l'ipotesi nulla H_0 del test bilaterale è falsa, il valore w osservato tende ad essere

maggiore o minore del valore atteso per la variabile W , in rapporto alla coda della distribuzione nella quale è collocata la tendenza centrale del gruppo con meno dati.

La potenza del test si esprime attraverso il P-valore, che nel test bilaterale rappresenta la probabilità che, sotto l'ipotesi H_0 , si osservi il valore w o uno più estremo cioè: se w è nella coda inferiore, P-valore= $2pr(w \geq W)$, se in quella superiore, P-valore= $2pr(w \leq W)$. Fissando la significatività del test a 0.05, nei casi in cui il P-valore risulta inferiore a questo valore l'ipotesi H_0 va rigettata.

Nel caso di test unilaterale sinistro calcoliamo il P-valore come la probabilità che sotto l'ipotesi di mediane uguali si osservi un valore W minore di w : P-valore= $pr(W < w)$, nel caso di test unilaterale destro P-valore= $pr(W > w)$.

Quando le dimensioni dei campioni sono maggiori di 20, possiamo approssimare la distribuzione di W con una distribuzione normale $\mathcal{N}(\mu, \sigma)$ dove, supponendo $n_A < n_B$, $\mu = \frac{n_A(n_A+n_B+1)}{2}$ e $\sigma = \sqrt{\frac{n_A n_B (n_A+n_B+1)}{12}}$.

Si approssima

$$pr(W \geq w) \approx pr(Z \geq z)$$

dove $z = \frac{w-\mu}{\sigma}$ e $Z = \mathcal{N}(0, 1)$ [29].

Appendice C

Tabelle

τ	(a) RRI			(b) SAP		
	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,887 ± 0,027	0,771 ± 0,020	0,0019	0,61 ± 0,03	0,851 ± 0,023	< 10 ⁻⁴
2	1,232 ± 0,026	1,054 ± 0,021	0,00016	0,98 ± 0,04	1,111 ± 0,023	0,021
3	1,345 ± 0,029	1,128 ± 0,020	< 10 ⁻⁵	1,15 ± 0,04	1,163 ± 0,022	0,72
4	1,391 ± 0,030	1,151 ± 0,019	< 10 ⁻⁷	1,22 ± 0,04	1,152 ± 0,020	0,18
5	1,41 ± 0,03	1,180 ± 0,017	< 10 ⁻⁷	1,27 ± 0,04	1,141 ± 0,018	0,0037
6	1,44 ± 0,03	1,23 ± 0,016	10 ⁻⁷	1,32 ± 0,04	1,175 ± 0,017	0,00050
7	1,45 ± 0,03	1,285 ± 0,016	< 10 ⁻⁴	1,36 ± 0,04	1,217 ± 0,016	0,00052
8	1,471 ± 0,029	1,333 ± 0,015	< 10 ⁻⁴	1,39 ± 0,03	1,267 ± 0,016	0,0021
9	1,50 ± 0,03	1,373 ± 0,016	0,0012	1,42 ± 0,03	1,327 ± 0,015	0,017
10	1,506 ± 0,026	1,417 ± 0,015	0,0092	1,46 ± 0,03	1,365 ± 0,015	0,018
τ	(c) DAP			(d) ILV		
	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,663 ± 0,028	0,936 ± 0,019	< 10 ⁻⁸	0,89 ± 0,03	0,886 ± 0,018	0,85
2	1,04 ± 0,03	1,212 ± 0,019	< 10 ⁻⁴	1,274 ± 0,028	1,188 ± 0,017	0,074
3	1,22 ± 0,03	1,246 ± 0,018	0,59	1,301 ± 0,026	1,211 ± 0,016	0,060
4	1,322 ± 0,029	1,244 ± 0,017	0,045	1,23 ± 0,03	1,135 ± 0,018	0,10
5	1,382 ± 0,028	1,244 ± 0,015	0,00018	1,14 ± 0,03	0,998 ± 0,019	0,0034
6	1,45 ± 0,03	1,284 ± 0,015	< 10 ⁻⁵	1,01 ± 0,04	0,922 ± 0,018	0,063
7	1,489 ± 0,027	1,335 ± 0,014	< 10 ⁻⁵	0,87 ± 0,05	0,908 ± 0,017	0,50
8	1,53 ± 0,03	1,372 ± 0,014	10 ⁻⁶	0,82 ± 0,04	0,921 ± 0,016	0,017
9	1,548 ± 0,028	1,420 ± 0,014	< 10 ⁻⁴	0,79 ± 0,04	0,940 ± 0,016	0,00055
10	1,58 ± 0,03	1,460 ± 0,014	< 10 ⁻⁴	0,80 ± 0,04	0,935 ± 0,016	0,0014

Tabella C.1: Valori medi delle entropie e relativi errori standard per i due gruppi in regime di respirazione basale, P-valori del WMW test

(a) RRI				(b) SAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,981 \pm 0,029	0,947 \pm 0,015	0,36	0,72 \pm 0,04	0,977 \pm 0,018	10 ⁻⁷
2	1,318 \pm 0,029	1,286 \pm 0,016	0,56	1,12 \pm 0,05	1,328 \pm 0,018	10 ⁻⁵
3	1,41 \pm 0,03	1,390 \pm 0,015	0,74	1,28 \pm 0,04	1,437 \pm 0,016	0,00026
4	1,45 \pm 0,03	1,427 \pm 0,015	0,71	1,33 \pm 0,04	1,465 \pm 0,016	0,00017
5	1,453 \pm 0,029	1,420 \pm 0,014	0,28	1,35 \pm 0,03	1,457 \pm 0,017	0,0045
6	1,45 \pm 0,03	1,405 \pm 0,014	0,10	1,37 \pm 0,03	1,398 \pm 0,017	0,32
7	1,445 \pm 0,027	1,382 \pm 0,013	0,028	1,39 \pm 0,03	1,347 \pm 0,016	0,32
8	1,49 \pm 0,03	1,409 \pm 0,014	0,012	1,45 \pm 0,03	1,335 \pm 0,017	0,0064
9	1,49 \pm 0,03	1,432 \pm 0,013	0,039	1,457 \pm 0,028	1,367 \pm 0,017	0,024
10	1,50 \pm 0,03	1,466 \pm 0,014	0,18	1,488 \pm 0,028	1,424 \pm 0,018	0,10

(c) DAP				(d) ILV		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,80 \pm 0,03	1,072 \pm 0,015	< 10 ⁻⁹	0,900 \pm 0,018	0,929 \pm 0,009	0,11
2	1,19 \pm 0,04	1,414 \pm 0,015	< 10 ⁻⁷	1,299 \pm 0,019	1,319 \pm 0,010	0,22
3	1,35 \pm 0,03	1,509 \pm 0,014	< 10 ⁻⁵	1,298 \pm 0,022	1,373 \pm 0,010	0,0014
4	1,427 \pm 0,028	1,521 \pm 0,014	0,0014	1,00 \pm 0,04	1,206 \pm 0,015	< 10 ⁻⁵
5	1,471 \pm 0,023	1,505 \pm 0,014	0,10	1,07 \pm 0,03	1,164 \pm 0,014	0,014
6	1,483 \pm 0,021	1,473 \pm 0,014	0,92	1,066 \pm 0,020	1,146 \pm 0,011	0,0016
7	1,522 \pm 0,020	1,435 \pm 0,014	0,0040	0,763 \pm 0,028	0,935 \pm 0,014	< 10 ⁻⁶
8	1,556 \pm 0,019	1,434 \pm 0,015	0,00011	0,52 \pm 0,04	0,718 \pm 0,016	< 10 ⁻⁵
9	1,565 \pm 0,020	1,472 \pm 0,015	0,0011	0,49 \pm 0,04	0,645 \pm 0,016	0,00019
10	1,567 \pm 0,022	1,506 \pm 0,016	0,028	0,57 \pm 0,03	0,687 \pm 0,016	0,0057

Tabella C.2: Valori medi delle entropie e relativi errori standard per i due gruppi in regime di respirazione controllata, P-valori del WMW test

(a) RRI → RRI				(b) SAP → RRI		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,936 ± 0,011	0,920 ± 0,007	0,89	0,003 ± 0,024	-0,004 ± 0,007	0,71
2	0,756 ± 0,014	0,776 ± 0,013	0,00082	0,03 ± 0,04	0,000 ± 0,011	0,75
3	0,656 ± 0,021	0,727 ± 0,014	< 10 ⁻⁴	0,02 ± 0,05	0,017 ± 0,012	0,16
4	0,587 ± 0,026	0,721 ± 0,013	< 10 ⁻⁶	0,01 ± 0,04	0,039 ± 0,011	0,037
5	0,550 ± 0,028	0,713 ± 0,012	< 10 ⁻⁷	0,00 ± 0,03	0,065 ± 0,010	0,018
6	0,51 ± 0,03	0,693 ± 0,012	< 10 ⁻⁷	-0,01 ± 0,03	0,086 ± 0,011	0,0025
7	0,49 ± 0,03	0,666 ± 0,012	< 10 ⁻⁷	-0,010 ± 0,028	0,102 ± 0,011	0,00092
8	0,45 ± 0,03	0,640 ± 0,012	10 ⁻⁸	-0,01 ± 0,03	0,112 ± 0,012	0,00026
9	0,42 ± 0,03	0,609 ± 0,013	< 10 ⁻⁷	-0,010 ± 0,028	0,121 ± 0,013	< 10 ⁻⁴
10	0,41 ± 0,03	0,581 ± 0,013	< 10 ⁻⁶	-0,009 ± 0,028	0,126 ± 0,014	< 10 ⁻⁴
(c) DAP → RRI				(d) ILV → RRI		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,061 ± 0,024	0,012 ± 0,005	< 10 ⁻⁴	0,145 ± 0,016	0,064 ± 0,009	< 10 ⁻⁵
2	0,04 ± 0,04	-0,028 ± 0,008	< 10 ⁻⁴	0,035 ± 0,004	0,0122 ± 0,0023	< 10 ⁻⁶
3	0,05 ± 0,05	-0,049 ± 0,009	< 10 ⁻⁶	0,018 ± 0,003	0,0051 ± 0,0015	< 10 ⁻⁵
4	0,067 ± 0,051	-0,049 ± 0,009	< 10 ⁻⁷	0,0119 ± 0,0025	0,0027 ± 0,0012	< 10 ⁻⁴
5	0,07 ± 0,04	-0,053 ± 0,008	< 10 ⁻⁷	0,0069 ± 0,0019	0,0015 ± 0,0011	0,00056
6	0,08 ± 0,04	-0,058 ± 0,009	< 10 ⁻⁸	0,0033 ± 0,0016	0,0008 ± 0,0010	0,084
7	0,069 ± 0,029	-0,065 ± 0,009	< 10 ⁻⁷	0,0003 ± 0,0015	0,0005 ± 0,0010	0,41
8	0,049 ± 0,028	-0,069 ± 0,009	< 10 ⁻⁷	-0,0015 ± 0,0016	0,0002 ± 0,0010	0,80
9	0,036 ± 0,024	-0,076 ± 0,010	< 10 ⁻⁷	-0,0023 ± 0,0015	0,0000 ± 0,0010	0,94
10	0,018 ± 0,021	-0,080 ± 0,010	< 10 ⁻⁷	-0,0027 ± 0,0015	0,0002 ± 0,0010	0,62

Tabella C.3: Valori medi dei coefficienti del VAR e relativi errori standard della regressione della serie RRI per i due gruppi in regime di respirazione basale, P-valori del WMW test

(a) RRI → SAP				(b) SAP → SAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	-0,061 ± 0,007	0,008 ± 0,006	10 ⁻⁸	0,894 ± 0,012	0,812 ± 0,012	0,0032
2	-0,195 ± 0,015	-0,118 ± 0,007	10 ⁻⁵	0,772 ± 0,026	0,683 ± 0,014	0,0087
3	-0,263 ± 0,019	-0,181 ± 0,008	< 10 ⁻⁴	0,67 ± 0,03	0,621 ± 0,015	0,26
4	-0,285 ± 0,021	-0,196 ± 0,008	< 10 ⁻⁴	0,59 ± 0,04	0,593 ± 0,016	0,93
5	-0,281 ± 0,022	-0,196 ± 0,008	0,00013	0,53 ± 0,05	0,575 ± 0,015	0,36
6	-0,273 ± 0,023	-0,194 ± 0,008	0,00065	0,50 ± 0,04	0,545 ± 0,015	0,27
7	-0,258 ± 0,022	-0,190 ± 0,009	0,0026	0,46 ± 0,04	0,510 ± 0,015	0,26
8	-0,242 ± 0,023	0,187 ± 0,010	0,014	0,45 ± 0,03	0,474 ± 0,015	0,35
9	-0,213 ± 0,023	-0,183 ± 0,010	0,16	0,42 ± 0,03	0,441 ± 0,016	0,45
10	-0,195 ± 0,025	-0,176 ± 0,011	0,32	0,39 ± 0,04	0,415 ± 0,016	0,46

(c) DAP → SAP				(d) ILV → SAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,068 ± 0,014	0,085 ± 0,008	0,76	0,149 ± 0,015	0,167 ± 0,009	0,50
2	0,104 ± 0,028	0,096 ± 0,011	0,52	0,060 ± 0,007	0,080 ± 0,004	0,11
3	0,135 ± 0,029	0,106 ± 0,011	0,13	0,032 ± 0,005	0,052 ± 0,004	0,019
4	0,15 ± 0,03	0,124 ± 0,011	0,35	0,020 ± 0,004	0,041 ± 0,004	0,0064
5	0,17 ± 0,04	0,139 ± 0,011	0,55	0,014 ± 0,004	0,037 ± 0,004	0,0011
6	0,16 ± 0,03	0,154 ± 0,011	0,77	0,011 ± 0,004	0,035 ± 0,004	0,00068
7	0,15 ± 0,03	0,168 ± 0,011	0,25	0,009 ± 0,004	0,034 ± 0,004	0,00023
8	0,136 ± 0,026	0,177 ± 0,012	0,10	0,010 ± 0,004	0,034 ± 0,004	0,00045
9	0,132 ± 0,028	0,184 ± 0,012	0,031	0,010 ± 0,004	0,033 ± 0,004	0,00074
10	0,130 ± 0,028	0,189 ± 0,013	0,017	0,011 ± 0,004	0,033 ± 0,004	0,008

Tabella C.4: Valori medi dei coefficienti del VAR e relativi errori standard della regressione della serie SAP per i due gruppi in regime di respirazione basale, P-valori del WMW test

(a) RRI → DAP				(b) SAP → DAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	-0,099 ± 0,007	-0,090 ± 0,006	0,15	-0,004 ± 0,017	0,105 ± 0,008	< 10 ⁻⁸
2	-0,200 ± 0,013	-0,179 ± 0,007	0,23	-0,040 ± 0,027	0,087 ± 0,010	< 10 ⁻⁷
3	-0,244 ± 0,017	-0,186 ± 0,007	0,0020	-0,113 ± 0,028	0,023 ± 0,011	10 ⁻⁷
4	-0,250 ± 0,019	-0,170 ± 0,007	< 10 ⁻⁴	-0,1767 ± 0,028	-0,034 ± 0,011	< 10 ⁻⁶
5	-0,239 ± 0,020	-0,151 ± 0,007	< 10 ⁻⁴	-0,23 ± 0,03	-0,075 ± 0,012	10 ⁻⁶
6	-0,216 ± 0,021	-0,135 ± 0,007	0,00021	-0,255 ± 0,029	-0,108 ± 0,012	10 ⁻⁵
7	-0,185 ± 0,020	-0,120 ± 0,008	0,0019	-0,28 ± 0,03	-0,143 ± 0,013	0,00013
8	-0,159 ± 0,021	-0,107 ± 0,008	0,018	-0,277 ± 0,029	-0,167 ± 0,014	0,00076
9	-0,123 ± 0,020	-0,095 ± 0,008	0,19	-0,28 ± 0,03	-0,183 ± 0,014	0,0062
10	-0,100 ± 0,020	-0,080 ± 0,008	0,34	-0,28 ± 0,03	-0,196 ± 0,015	0,017

(c) DAP → DAP				(d) ILV → DAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,947 ± 0,016	0,746 ± 0,010	< 10 ⁻¹⁵	0,109 ± 0,013	0,154 ± 0,008	0,014
2	0,889 ± 0,028	0,626 ± 0,013	< 10 ⁻¹³	0,074 ± 0,008	0,104 ± 0,005	0,0082
3	0,87 ± 0,03	0,644 ± 0,015	10 ⁻¹⁰	0,060 ± 0,008	0,084 ± 0,004	0,025
4	0,86 ± 0,03	0,691 ± 0,014	< 10 ⁻⁶	0,054 ± 0,008	0,077 ± 0,004	0,055
5	0,85 ± 0,04	0,716 ± 0,015	0,00020	0,052 ± 0,008	0,074 ± 0,005	0,068
6	0,82 ± 0,03	0,726 ± 0,015	0,014	0,052 ± 0,008	0,074 ± 0,005	0,084
7	0,79 ± 0,03	0,738 ± 0,015	0,19	0,052 ± 0,008	0,074 ± 0,005	0,092
8	0,76 ± 0,03	0,734 ± 0,015	0,48	0,053 ± 0,008	0,074 ± 0,005	0,099
9	0,74 ± 0,04	0,730 ± 0,015	0,96	0,054 ± 0,008	0,075 ± 0,005	0,11
10	0,72 ± 0,04	0,728 ± 0,015	0,73	0,056 ± 0,008	0,075 ± 0,005	0,13

Tabella C.5: Valori medi dei coefficienti del VAR e relativi errori standard della regressione della serie DAP per i due gruppi in regime di respirazione basale, P-valori del WMW test

(a) RRI → ILV				(b) SAP → ILV		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	-0,137 ± 0,016	-0,148 ± 0,010	0,98	-0,15 ± 0,04	-0,108 ± 0,012	0,52
2	-0,024 ± 0,006	-0,034 ± 0,003	0,23	-0,012 ± 0,015	-0,030 ± 0,007	0,27
3	-0,005 ± 0,004	-0,0162 ± 0,0022	0,012	0,025 ± 0,011	0,002 ± 0,006	0,061
4	0,001 ± 0,003	-0,0114 ± 0,0018	0,0010	0,033 ± 0,009	0,012 ± 0,005	0,037
5	0,0001 ± 0,0026	-0,010 ± 0,0018	0,016	0,033 ± 0,008	0,012 ± 0,005	0,026
6	-0,0015 ± 0,0023	-0,0086 ± 0,0017	0,12	0,030 ± 0,007	0,011 ± 0,005	0,033
7	-0,0026 ± 0,0021	-0,0077 ± 0,0016	0,12	0,027 ± 0,006	0,010 ± 0,005	0,037
8	-0,0034 ± 0,0019	-0,0065 ± 0,0016	0,35	0,023 ± 0,005	0,008 ± 0,005	0,0632
9	-0,0028 ± 0,0018	-0,0052 ± 0,0016	0,44	0,021 ± 0,005	0,008 ± 0,005	0,091
10	-0,0027 ± 0,0016	-0,0046 ± 0,0016	0,57	0,019 ± 0,005	0,007 ± 0,005	0,13

(c) DAP → ILV				(d) ILV → ILV		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,01 ± 0,04	-0,085 ± 0,012	0,010	0,694 ± 0,022	0,685 ± 0,010	1,0
2	-0,060 ± 0,018	-0,064 ± 0,008	0,77	0,722 ± 0,017	0,705 ± 0,010	0,86
3	-0,070 ± 0,014	-0,066 ± 0,007	0,60	0,726 ± 0,017	0,708 ± 0,010	0,85
4	-0,068 ± 0,013	-0,067 ± 0,007	0,63	0,726 ± 0,017	0,709 ± 0,010	0,88
5	-0,064 ± 0,012	-0,065 ± 0,007	0,621	0,727 ± 0,017	0,709 ± 0,010	0,87
6	-0,060 ± 0,011	-0,064 ± 0,007	0,70569	0,727 ± 0,017	0,710 ± 0,010	0,86
7	-0,057 ± 0,011	-0,064 ± 0,007	0,79	0,728 ± 0,017	0,710 ± 0,010	0,85
8	-0,054 ± 0,011	-0,063 ± 0,006	0,90	0,728 ± 0,017	0,710 ± 0,010	0,84
9	-0,052 ± 0,011	-0,063 ± 0,006	0,981	0,728 ± 0,017	0,710 ± 0,010	0,83
10	-0,052 ± 0,011	-0,063 ± 0,006	0,884	0,728 ± 0,017	0,710 ± 0,010	0,85

Tabella C.6: Valori medi dei coefficienti del VAR e relativi errori standard della regressione della serie ILV per i due gruppi in regime di respirazione basale, P-valori del WMW test

(a) RRI → RRI				(b) SAP → RRI		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,915 ± 0,015	0,886 ± 0,008	0,38	-0,008 ± 0,022	0,002 ± 0,009	0,66
2	0,687 ± 0,020	0,686 ± 0,014	0,26	0,05 ± 0,04	-0,009 ± 0,015	0,15
3	0,57 ± 0,03	0,578 ± 0,017	0,37	0,02 ± 0,04	-0,038 ± 0,017	0,20
4	0,51 ± 0,04	0,522 ± 0,019	0,27	-0,03 ± 0,03	-0,049 ± 0,015	0,60
5	0,50 ± 0,03	0,536 ± 0,017	0,12	-0,039 ± 0,027	-0,032 ± 0,012	0,72
6	0,51 ± 0,03	0,578 ± 0,013	0,031	-0,022 ± 0,026	-0,002 ± 0,010	0,22
7	0,49 ± 0,03	0,593 ± 0,011	0,0041	0,000 ± 0,028	0,026 ± 0,009	0,13
8	0,45 ± 0,04	0,580 ± 0,010	0,00065	-0,013 ± 0,027	0,040 ± 0,009	0,0071
9	0,42 ± 0,04	0,551 ± 0,011	0,0028	-0,015 ± 0,027	0,041 ± 0,009	0,0044
10	0,39 ± 0,04	0,520 ± 0,011	0,0026	-0,032 ± 0,029	0,042 ± 0,009	0,00099
(c) DAP → RRI				(d) ILV → RRI		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,098 ± 0,019	0,028 ± 0,006	< 10 ⁻⁴	0,159 ± 0,019	0,112 ± 0,012	0,091
2	0,06 ± 0,03	0,010 ± 0,011	0,073	0,0471 ± 0,0059	0,039 ± 0,005	0,13
3	0,08 ± 0,04	0,002 ± 0,012	0,018	0,0280 ± 0,0037	0,0203 ± 0,0029	0,033
4	0,10 ± 0,03	-0,004 ± 0,010	0,00027	0,0179 ± 0,0027	0,0105 ± 0,0022	0,015
5	0,112 ± 0,028	-0,001 ± 0,008	10 ⁻⁵	0,0089 ± 0,0020	0,0041 ± 0,0017	0,022
6	0,108 ± 0,026	0,010 ± 0,007	< 10 ⁻⁵	0,0030 ± 0,0016	0,0008 ± 0,0012	0,16
7	0,092 ± 0,024	0,006 ± 0,006	< 10 ⁻⁴	0,0013 ± 0,0011	0,0000 ± 0,0009	0,27
8	0,076 ± 0,022	0,007 ± 0,006	0,00031	0,0005 ± 0,0009	0,0003 ± 0,0008	0,51
9	0,059 ± 0,020	0,004 ± 0,006	0,0012	0,0002 ± 0,0009	0,0002 ± 0,0008	0,69
10	0,048 ± 0,017	0,0025 ± 0,005	0,0034	0,0005 ± 0,0009	0,0002 ± 0,0008	0,32

Tabella C.7: Valori medi dei coefficienti del VAR e relativi errori standard della regressione della serie RRI per i due gruppi in regime di respirazione controllata, P-valori del WMW test

(a) RRI → SAP				(b) SAP → SAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	-0,072 ± 0,008	0,007 ± 0,005	10 ⁻¹⁰	0,897 ± 0,017	0,823 ± 0,012	0,0090
2	-0,216 ± 0,017	-0,122 ± 0,009	< 10 ⁻⁵	0,75 ± 0,03	0,646 ± 0,015	0,0016
3	-0,277 ± 0,018	-0,197 ± 0,011	0,00077	0,58 ± 0,05	0,495 ± 0,020	0,020
4	-0,282 ± 0,020	-0,213 ± 0,011	0,0079	0,48 ± 0,06	0,412 ± 0,023	0,10
5	-0,262 ± 0,022	-0,193 ± 0,010	0,0047	0,46 ± 0,05	0,422 ± 0,021	0,32
6	-0,245 ± 0,023	-0,163 ± 0,009	0,00076	0,47 ± 0,04	0,458 ± 0,017	0,67
7	-0,234 ± 0,022	-0,142 ± 0,008	0,00010	0,45 ± 0,04	0,473 ± 0,016	0,53
8	-0,222 ± 0,023	-0,131 ± 0,008	0,00016	0,41 ± 0,04	0,452 ± 0,017	0,27
9	-0,201 ± 0,024	-0,123 ± 0,009	0,0026	0,37 ± 0,04	0,420 ± 0,017	0,23
10	-0,174 ± 0,024	-0,111 ± 0,009	0,017	0,33 ± 0,05	0,390 ± 0,017	0,14

(c) DAP → SAP				(d) ILV → SAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,052 ± 0,020	0,093 ± 0,009	0,36	0,195 ± 0,019	0,256 ± 0,010	0,0028
2	0,08 ± 0,04	0,084 ± 0,012	0,93	0,094 ± 0,011	0,141 ± 0,007	0,00038
3	0,13 ± 0,04	0,094 ± 0,014	0,40	0,056 ± 0,008	0,095 ± 0,005	0,00019
4	0,18 ± 0,04	0,117 ± 0,013	0,20	0,036 ± 0,006	0,072 ± 0,005	< 10 ⁻⁴
5	0,18 ± 0,04	0,130 ± 0,012	0,26	0,025 ± 0,005	0,060 ± 0,005	< 10 ⁻⁴
6	0,16 ± 0,04	0,129 ± 0,011	0,80	0,020 ± 0,004	0,056 ± 0,005	< 10 ⁻⁴
7	0,14 ± 0,04	0,125 ± 0,010	0,67	0,019 ± 0,004	0,058 ± 0,005	0,00010
8	0,13 ± 0,04	0,129 ± 0,011	0,36	0,020 ± 0,004	0,059 ± 0,005	0,00013
9	0,13 ± 0,04	0,134 ± 0,011	0,16	0,021 ± 0,005	0,060 ± 0,005	0,00015
10	0,13 ± 0,04	0,138 ± 0,011	0,13	0,021 ± 0,005	0,061 ± 0,005	0,00020

Tabella C.8: Valori medi dei coefficienti del VAR e relativi errori standard della regressione della serie SAP per i due gruppi in regime di respirazione controllata, P-valori del WMW test

(a) RRI → DAP				(b) SAP → DAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	-0,113 ± 0,009	-0,088 ± 0,005	0,012	0,032 ± 0,022	0,144 ± 0,009	< 10 ⁻⁷
2	-0,227 ± 0,016	-0,168 ± 0,009	0,0052	0,00 ± 0,04	0,118 ± 0,013	< 10 ⁻⁴
3	-0,264 ± 0,017	-0,184 ± 0,010	0,00046	-0,11 ± 0,04	0,014 ± 0,015	0,0010
4	-0,250 ± 0,018	-0,166 ± 0,009	0,00023	-0,21 ± 0,04	-0,062 ± 0,015	0,00029
5	-0,220 ± 0,019	-0,133 ± 0,008	< 10 ⁻⁴	-0,25 ± 0,04	-0,085 ± 0,014	< 10 ⁻⁵
6	-0,186 ± 0,019	-0,101 ± 0,006	< 10 ⁻⁴	-0,26 ± 0,03	-0,083 ± 0,013	< 10 ⁻⁶
7	-0,156 ± 0,019	-0,080 ± 0,006	< 10 ⁻⁴	-0,27 ± 0,03	-0,085 ± 0,013	< 10 ⁻⁶
8	-0,133 ± 0,017	-0,065 ± 0,006	0,00022	-0,28 ± 0,03	-0,099 ± 0,014	< 10 ⁻⁶
9	-0,107 ± 0,016	-0,054 ± 0,006	0,0054	-0,29 ± 0,03	-0,113 ± 0,014	< 10 ⁻⁵
10	-0,079 ± 0,015	-0,042 ± 0,006	0,088	-0,30 ± 0,04	-0,126 ± 0,014	< 10 ⁻⁵

(c) DAP → DAP				(d) ILV → DAP		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	0,917 ± 0,022	0,745 ± 0,009	10 ⁻¹³	0,165 ± 0,017	0,217 ± 0,008	0,0091
2	0,82 ± 0,04	0,585 ± 0,014	< 10 ⁻¹⁰	0,121 ± 0,011	0,151 ± 0,006	0,044
3	0,82 ± 0,05	0,567 ± 0,016	< 10 ⁻⁸	0,105 ± 0,010	0,126 ± 0,005	0,12
4	0,82 ± 0,04	0,591 ± 0,016	< 10 ⁻⁸	0,096 ± 0,010	0,115 ± 0,005	0,18
5	0,83 ± 0,04	0,618 ± 0,015	< 10 ⁻⁷	0,091 ± 0,010	0,111 ± 0,005	0,18
6	0,80 ± 0,03	0,628 ± 0,015	< 10 ⁻⁵	0,089 ± 0,010	0,112 ± 0,005	0,13
7	0,78 ± 0,04	0,628 ± 0,015	< 10 ⁻⁴	0,090 ± 0,010	0,114 ± 0,006	0,11
8	0,75 ± 0,04	0,625 ± 0,015	0,00094	0,091 ± 0,010	0,115 ± 0,006	0,10
9	0,73 ± 0,04	0,621 ± 0,015	0,0056	0,091 ± 0,011	0,117 ± 0,006	0,10
10	0,72 ± 0,04	0,618 ± 0,015	0,012	0,092 ± 0,011	0,117 ± 0,006	0,096

Tabella C.9: Valori medi dei coefficienti del VAR e relativi errori standard della regressione della serie DAP per i due gruppi in regime di respirazione controllata, P-valori del WMW test

(a) RRI → ILV				(b) SAP → ILV		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	-0,151 ± 0,021	-0,122 ± 0,010	0,33	-0,15 ± 0,05	-0,145 ± 0,015	0,84
2	-0,019 ± 0,008	-0,033 ± 0,005	0,12	-0,024 ± 0,021	-0,043 ± 0,011	0,60
3	0,006 ± 0,006	-0,016 ± 0,004	0,0023	0,035 ± 0,015	0,016 ± 0,009	0,48
4	0,013 ± 0,005	-0,009 ± 0,003	0,00034	0,059 ± 0,012	0,044 ± 0,008	0,64
5	0,009 ± 0,003	-0,0066 ± 0,0024	0,00087	0,069 ± 0,011	0,051 ± 0,007	0,50
6	0,0029 ± 0,0024	-0,0083 ± 0,0018	0,0074	0,066 ± 0,009	0,047 ± 0,007	0,42
7	-0,0019 ± 0,0019	-0,0102 ± 0,0015	0,012	0,058 ± 0,009	0,037 ± 0,007	0,31
8	-0,0026 ± 0,0017	-0,0112 ± 0,0014	0,0027	0,054 ± 0,008	0,031 ± 0,006	0,23
9	-0,0020 ± 0,0016	-0,0097 ± 0,0013	0,0025	0,050 ± 0,008	0,027 ± 0,006	0,26
10	-0,0017 ± 0,0015	-0,0094 ± 0,0013	0,0015	0,047 ± 0,008	0,025 ± 0,006	0,30
(c) DAP → ILV				(d) ILV → ILV		
τ	Sani	Malati	P-valori	Sani	Malati	P-valori
1	-0,07 ± 0,05	-0,131 ± 0,012	0,047	0,677 ± 0,014	0,716 ± 0,007	0,0084
2	-0,110 ± 0,023	-0,138 ± 0,010	0,14	0,719 ± 0,007	0,730 ± 0,005	0,091
3	-0,122 ± 0,018	-0,144 ± 0,009	0,22	0,722 ± 0,007	0,735 ± 0,005	0,031
4	-0,122 ± 0,016	-0,146 ± 0,009	0,19	0,722 ± 0,007	0,736 ± 0,005	0,026
5	-0,122 ± 0,015	-0,144 ± 0,008	0,25	0,724 ± 0,007	0,738 ± 0,004	0,024
6	-0,120 ± 0,014	-0,141 ± 0,008	0,24	0,726 ± 0,007	0,740 ± 0,005	0,025
7	-0,116 ± 0,014	-0,138 ± 0,008	0,22	0,727 ± 0,007	0,741 ± 0,004	0,022
8	-0,113 ± 0,014	-0,135 ± 0,008	0,22	0,727 ± 0,007	0,741 ± 0,004	0,021
9	-0,110 ± 0,014	-0,133 ± 0,008	0,19	0,727 ± 0,007	0,741 ± 0,004	0,021
10	-0,108 ± 0,014	-0,132 ± 0,008	0,17	0,727 ± 0,007	0,741 ± 0,004	0,020

Tabella C.10: Valori medi dei coefficienti del VAR e relativi errori standard della regressione della serie ILV per i due gruppi in regime di respirazione controllata, P-valori del WMW test

Bibliografia

- [1] L. Devroye, L. Györfi, G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, (Springer-Verlag, New York, 1996).
- [2] P. van Leeuwen, H. Bettermann, *The status of nonlinear dynamics in the analysis of heart rate variability*, *Herzschr Elektrophys*,bf 11, 127-130 (Steinkopff Verlag, 2000)
- [3] D. P. Feldman, *Complexity Measure Discussion*, (SFI CSSS, Beijing China, July 2005 da <http://www.hornacek.coa.edu/dave>)
- [4] R. Badii, A. Politi, *Complexity, Hierarchical structures and scaling in physics*, (Cambridge Nonlinear Science Series, vol.6, Cambridge University Press, 1997).
- [5] H. Kantz, T. Schreiber, *Nonlinear time series analysis*, (Cambridge Nonlinear Science Series, vol.7, Cambridge University Press, 1997).
- [6] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, *Gene Selection for Cancer Classification using Support Vector Machines*, *Machine Learning*, **46**, 389-422 (2002).
- [7] S. Guzzetti, *Variabilità della frequenza cardiaca*, *Ital Heart J Suppl*, **2**(5),450-454, (2001)

- [8] M. Costa, A. L. Goldberger, C.-K. Peng, *Multiscale entropy analysis of biological signals*, Physical Review E **71**, 021906 (2005)
- [9] L. Mangin, A. Monti, C. Médigue, I. Macquin-Mavier, M.-E. Lopes, P. Gueret, A. Castaigne, B. Swynghedauw, P. Mansier, *Altered baroreflex gain during voluntary breathing in chronic heart failure*, European Journal of Heart Failure, **3**, 189-195, (2001)
- [10] T. Schreiber, *Interdisciplinary application of nonlinear time series methods*, Physics Reports **308**, Elsevier Science (1999)
- [11] Rajendra Acharya U*, Kannathal N, Ong Wai Sing, Luk Yi Ping and Tji-Leng Chua, *Heart rate analysis in normal subjects of various age groups*, (BioMedical Engineering OnLine 2004, da <http://www.biomedical-engineering-online.com/content/3/1/24>)
- [12] E. Ott, *Chaos in Dynamical Systems*, (Cambridge University Press, 2002)
- [13] A. Cohen, I. Procaccia, *Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems*, Physical Review A **31**, 1872 (1985)
- [14] S. M. Pincus, A. L. Goldberger, *Physiological time-series analysis: what does regularity quantify?*, Am. J. Physiol.266 (Heart circ. Physiol. **35**), H1643-H1656, (1994)
- [15] M. Costa, A. L. Goldberger, C.-K. Peng, *Multiscale entropy analysis of complex physiologic time series*, Physical Review Letters **89**, 068102 (2002)
- [16] J. P. Eckmann, D. Ruelle, *Ergodic theory of chaos and strange attractors*, Reviews of Modern Physics, **57**, 617, (1985)

- [17] S. M. Pincus, *Approximate entropy as a measure of system complexity*, Proc. Natl. Acad. Sci. USA, **77**, 2297 (1991)
- [18] J. S. Richmann, J. R. Moorman, *Physiological time-series analysis using approximate entropy and sample entropy*, Am J Physiol Heart Circ Physiol, **278**, H2039, (2000)
- [19] J. D. Hamilton, *Time Series Analysis*, (Princeton University Press, 1994)
- [20] C. Prati, *Segnali e sistemi per le telecomunicazioni*, (McGraw-Hill Milano, 2003)
- [21] V. N. Vapnik, *Statistical Learning Theory*, (Wiley New York, 1998)
- [22] T. W. Anderson, *An introduction to multivariate statistical analysis*, (New York: John Wiley, 1958)
- [23] W. J. Krzanowski, *Principles of multivariate analysis: a user's perspective*, (Oxford Science Publications, 1988)
- [24] S. Mukherjee, R. Rifkin, T. Poggio, *Regression and classification with regularization*, Lectures Notes in Statistics: Nonlinear Estimation and Classification, Proc. MSRI Workshop, **171** ed D. D. Denison, M. H. Hansen, C. C. Holmes, B. Yu, 107-124 (Berlin: Springer, 2002)
- [25] N. Ancona, R. Maestri, D. Marinazzo, L. Nitti, M. Pellicoro, G. D. Pinna, S. Stramaglia, *Physiol. Meas.*, **26**, 363-372, (2005)
- [26] P. Wahl, R. A. Kronmal, *Discriminant functions when covariances are unequal and sample sizes are moderate*, Biometrics, **33**, 479-484, (1977)
- [27] E. Bottarelli, S. Parodi, *Un approccio per la valutazione della validità dei test diagnostici: le curve R.O.C. (Receiver Operating Characteristic)*, Ann. Fac. Medic. Vet. di Parma (Vol. XXIII, 2003)

[28] <http://www.dsa.unipr.it/soliani/capu9.pdf>

[29] <http://www.stat.auckland.ac.nz/~wild/ChanceEnc/Ch10.wilcoxon.pdf>

Ringraziamenti

Ringrazio il Prof. Pellicoro per avermi guidato sapientemente in questo lavoro di tesi attraverso utili discussioni e chiarimenti.

Voglio ringraziare il Prof. Stramaglia per avermi aiutato con i suoi insegnamenti di fisica uniti sempre a rassicurazioni e incoraggiamenti.

Un grazie a Daniele che mi ha introdotto all'analisi dei segnali fisiologici con tutto il suo entusiasmo.

Devo ringraziare inoltre il Prof. Angelini per avermi dedicato tempo prezioso (com'è solitamente quello della pausa pranzo) per mettere a disposizione le sue conoscenze attraverso un dialogo aperto e stimolante.

Ringrazio Rossella e Lorenzo che hanno preso a cuore la buona riuscita di questo lavoro e il mio stato d'animo, soprattutto nei momenti più difficili, con un interesse spropositato.

Un grande grazie a Vito, Vittoria, Giuseppe e Lucia per essermi stati sempre amorevolmente vicini.

Ringrazio anche Mino e Lucia per aver rinunciato a vedere la semifinale dei mondiali di calcio per problemi tecnici legati alla stampa della mia tesi.

Un sincero grazie a Donato per la sua estrema disponibilità e infine a tutti quelli che mi hanno sostenuto e incoraggiato durante l'intero corso di laurea.