



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA BIOMEDICA

ALGORITMI BASATI SU SUPPORT VECTOR REGRESSION

PER LA PREDIZIONE DELLA GLICEMIA FUTURA

NELLA TERAPIA DEL DIABETE DI TIPO 1

Relatore: Prof. Andrea Facchinetti

Laureanda: Giada Businaro

ANNO ACCADEMICO 2021-2022

Data di laurea 20 Settembre

RINGRAZIAMENTI

Prima di procedere con la trattazione, volevo ringraziare le persone che mi hanno aiutato a raggiungere questo traguardo.

Ringrazio il mio relatore per la sua disponibilità e le conoscenze trasmesse durante la stesura dell'elaborato.

Un grazie di cuore ai miei genitori, che hanno sempre creduto in me e non mi hanno mai fatto perdere le speranze.

SOMMARIO

Il diabete mellito è una patologia epidemica mondiale e, secondo i più recenti dati dell'Organizzazione Mondiale della Sanità, si classifica come la nona causa di morte nel mondo. Si distinguono due diverse tipologie di diabete mellito: il diabete di tipo 1 e il diabete di tipo 2. Il primo solitamente insorge durante l'infanzia o l'adolescenza e le cause sono su base immunitaria; il secondo compare nella maggior parte in pazienti di età adulta, spesso collegato all'obesità ed a uno stile di vita sedentario. Per i pazienti diabetici è necessario mantenere il livello di glucosio all'interno di un determinato range, per prevenire episodi di ipo ed iperglicemia. Per questo monitorare i livelli di glucosio ha portato notevoli vantaggi ai pazienti, consentendo loro di prendere il controllo sulla malattia e ridurre l'incidenza di complicanze. Lo scopo di questa tesi è descrivere il machine learning, ponendo la nostra attenzione negli algoritmi di support vector regression (SVR).

Il primo capitolo è dedicato agli aspetti generali della patologia del diabete e presenta una panoramica dei dispositivi per rilevare la concentrazione di glucosio nel sangue. Questi ultimi sono potenzialmente utili perché, dai dati da essi raccolti, si possono generare allarmi quando la concentrazione di glucosio va al di sopra o al di sotto di una certa soglia.

Nel secondo capitolo sono analizzate le principali caratteristiche delle support vector machine e la loro principale distinzione: classificazione e regressione. Ci focalizzeremo sulla seconda, analizzando la realizzazione e il funzionamento dell'algoritmo SVR.

Il terzo capitolo è interamente occupato da un caso di studio, in cui viene predetto il segnale glicemico tramite l'utilizzo dell'algoritmo SVR. Lo scopo del capitolo è mostrare come viene in pratica utilizzato SVR, e come le predizioni ottenute siano diverse a seconda dei dati di ingresso a disposizione e alla scelta di determinati parametri.

Infine vengono fatte alcune considerazioni conclusive sui risultati del precedente studio e sull'utilità degli algoritmi che generano allarmi futuri per il segnale glicemico. Viene dedicato, inoltre, un piccolo spazio per le prospettive future.

INDICE

CAPITOLO 1: IL DIABETE	7
1.1 Aspetti generali del diabete mellito	7
1.2 Il diabete di tipo 1	7
1.3 Il diabete di tipo 2	8
1.4 Monitoraggio del segnale glicemico	9
1.5 Sensori per monitorare la glicemia	11
1.5.1 Biosensori minimamente invasivi	11
1.5.2 Biosensori non invasivi	14
Capitolo 2: LA SUPPORT VECTOR REGRESSION	16
2.1 Machine learning per la predizione del glucosio nel sangue	17
2.2 Concetti generali sulle support vector machines	18
2.3 Scopo delle support vector machines	19
2.4 Cenni su support vector classifier (SVC)	20
2.4.1 SVC dal punto di vista geometrico	20
2.4.2 Principali proprietà di SVC	21
2.4.3 Controllo della dimensione di Vapnik-Chervonenkis (VC)	22
2.4.2 Hard Margin e Soft margin	23
2.5 Support vector regression (SVR)	24
2.5.1 Principi di funzionamento	24
2.5.2 Significato del parametro w dell'iperpiano di separazione	26
2.5.3 Modello matematico per la regressione lineare	27
2.5.4 Funzione di perdita	28
2.5.5 Introduzione delle variabili slack	29
2.6 Kernel	30
2.7 Rischio funzionale	31
CAPITOLO 3: APPLICAZIONE DI SVR PER LA PREDIZIONE DEL SEGNALE DI GLUCOSIO IN CONTINUA	33

3.1 Caso di studio: predizione multivariata della concentrazione di glucosio sottocutaneo in pazienti affetti da diabete di tipo 1 basata su Support Vector Regression	33
3.1.1 Dati	33
3.1.2 Algoritmo	35
3.1.3 Valutazione del modello	36
3.1.4 Ottimizzazione di parametri	37
3.1.5 Risultati dell'applicazione	37
CONCLUSIONI	41
BIBLIOGRAFIA	42

CAPITOLO 1: IL DIABETE

1.1 Aspetti generali del diabete mellito

Il diabete mellito non è una sola malattia, ma un gruppo eterogeneo di sindromi multifattoriali con origini diverse e caratterizzate da un innalzamento del livello del glucosio ematico a digiuno per una carenza relativa o assoluta di insulina. Gli aggiustamenti che istante per istante mantengono i livelli di glucosio a tra i 70 e 180 mg/dl (range euglicemico) coinvolgono l'azione coordinata di due principali ormoni, l'insulina e il glucagone. L'insulina viene prodotta dalle cellule β degli isolotti di Langerhans, raggruppamenti di cellule presenti nel pancreas. L'insulina segnala ai tessuti che la concentrazione di glucosio nel sangue è più alta del necessario, e fa in modo che l'eccesso di glucosio venga assorbito dal sangue e convertito in altri composti di riserva, come il glicogeno e i trigliceridi. Il glucagone trasmette invece il messaggio che la disponibilità di glucosio è limitata, per cui i tessuti rispondono producendo glucosio attraverso la gluconeogenesi (nel fegato). [1]

Secondo l'ultimo rapporto fornito dall'Organizzazione Mondiale della Sanità, il numero di persone con diabete è passato da 108 milioni nel 1980 a 422 milioni nel 2014, che significa che l'8,5% delle persone di età superiore a 18 anni possiede questa malattia, con una stima di più di 500 milioni di casi entro il 2030. Il diabete è la causa principale di cecità e di amputazione nell'adulto, ma interviene anche nei casi di insufficienza renale, danni nervosi, attacchi cardiaci e ictus.[2]

Il diabete può essere classificato in due gruppi: il tipo 1(T1D), denominato diabete mellito insulina-dipendente e il tipo 2(T2D), denominato diabete-mellito insulina-indipendente.[3]

1.2 Il diabete di tipo 1

La malattia è caratterizzata da una totale mancanza di insulina, dovuta a un'aggressione autoimmune delle cellule β del pancreas. Ciò comporta un graduale impoverimento della popolazione di cellule β , ma i sintomi compaiono soltanto quando la percentuale di cellule distrutte è tra l'80 e il 90%. Come conseguenza il pancreas non riesce a rispondere in modo adeguato all'assunzione di glucosio ed è necessaria la somministrazione di insulina esogena per ripristinare il controllo metabolico.

Sono due le possibili cause della distruzione delle cellule β ; uno stimolo proveniente dall'esterno, per esempio un'infezione virale, oppure un determinante genico, che porta a

riconoscere le cellule come “estrane”. L’esordio del diabete di tipo 1 avviene tipicamente nell’infanzia o alla pubertà. Generalmente i sintomi sono poliuria (orinazione frequente), polifagia (fame eccessiva) e polidipsia (sete eccessiva), accompagnati da affaticamento e perdita di peso.

I tessuti più colpiti dalle alterazioni dei livelli di insulina e glucagone sono il fegato, i muscoli e il tessuto adiposo. Il segno caratteristico del diabete di tipo 1 non trattato è un elevato livello ematico di glucosio, causato dall’ aumento della produzione epatica di glucosio per attivazione della via gluconeogenica e dalla diminuzione dell’assorbimento del glucosio da parte delle cellule del tessuto muscolare e adiposo.

I trattamenti ricorrenti per trattare il diabete di tipo 1 sono due: il trattamento standard con insulina e il trattamento invasivo con insulina. Il primo consiste nei casi tipici in una o due somministrazioni al giorno di insulina, ottenendo valori di glucosio che variano nell’intervallo compreso tra i 225 e i 275 mg/dl. Il secondo cerca di rendere il range euglicemico più simile a quello di un individuo sano, tramite un monitoraggio più frequente, somministrando insulina per tre o più volte al giorno; infatti il valore di glicemia raggiunto con quest’ultimo trattamento è circa 150 mg/dl. Il trattamento intensivo si è rivelato migliore rispetto a quello standard perché riduce del 50% o più complicanze a lungo termine della malattia (retinopatia, nefropatia). Tuttavia è importante raggiungere anche il giusto dosaggio di insulina, tenendo presente che l’ipoglicemia è la complicanza più frequente nel diabete di tipo 1. Il trattamento intensivo comporta una maggiore frequenza di episodi ipoglicemici in confronto al trattamento standard.

1.3 Il diabete di tipo 2

L’incidenza del diabete di tipo 2 è in considerevole aumento a causa dell’invecchiamento della popolazione e ad uno stile di vita sedentario. La malattia insorge con gradualità ed è rilevabile da esami del sangue di routine. Sintomi evidenti possono essere poliuria, polidipsia, talvolta polifagia. Il diabete di tipo 2 è la combinazione di resistenza all’insulina e malfunzionamento delle cellule, ed entrambi portano all’iperglicemia. La patogenesi, a differenza del diabete di tipo 1, non coinvolge virus o anticorpi autoimmuni e non è ancora stata completamente compresa. La resistenza all’insulina di solito si verifica negli anziani, negli individui obesi o fisicamente inattivi e in piccola percentuale, in donne in gravidanza con diabete gestazionale. L’accumulo di grasso in eccesso nel tessuto adiposo è proporzionale allo sviluppo di una resistenza all’insulina.

Per trattare il diabete di tipo 2, si cerca di mantenere la concentrazione ematica del glucosio entro i limiti standard e di prevenire le complicanze a lungo termine della malattia. Per la maggior parte delle persone affette da questa malattia, è sufficiente la riduzione del peso corporeo, l'esercizio fisico e la dieta. In altri casi vengono utilizzati anche agenti ipoglicemizzanti, per esempio la metformina (che diminuisce la secrezione di glucosio da parte del fegato), le solfoniluree (che aumentano la secrezione di insulina), gli inibitori della α -glucosidasi (che diminuiscono l'assorbimento dei carboidrati introdotti con l'alimentazione) o la terapia insulinica.[3]

1.4 Monitoraggio del segnale glicemico

La conoscenza della concentrazione di glucosio nel sangue è un aspetto chiave nella comprensione quantitativa del segnale glicemico, nella diagnosi e nel trattamento del diabete. Alcuni decenni fa, negli anni '80 e '90, sono iniziati gli studi sulle tecniche di elaborazione del segnale di glucosio, ottenendo serie temporali del segnale da campioni in quantità sufficientemente alti di sangue. Venivano utilizzati metodi lineari, come analisi di correlazione, spettro, rilevamento del picco e metodi non lineari. Questi studi si basavano sull'automonitoraggio della glicemia (SMBG), il quale è considerato tutt'ora da parte dell'Associazione Diabetica Americana (ADA) parte integrante di qualsiasi strategia di gestione diretta al raggiungimento della normalizzazione della glicemia. Per la maggior parte dei pazienti di tipo 1, SMBG è raccomandato tre o più volte al giorno, anche se la frequenza può variare in base all'instabilità del segnale glicemico. Per quanto riguarda il diabete di tipo 2, le idee sono controverse riguardo l'utilità dell'SMBG, tuttavia si è d'accordo sul fatto che possa essere utile in presenza di malattie ricorrenti, modifiche della terapia o situazioni di maggior rischio.[4]

I dispositivi per l'automonitoraggio del glucosio sono i glucometri; sono costituiti da un ago che consente il prelievo di una goccia di sangue capillare da un polpastrello. Il glucosio contenuto nella goccia di sangue dà origine a una reazione enzimatica ossidandosi a gluconolattone grazie alla presenza dell'enzima glucosio-ossidasi; tale variazione darà origine a una variazione cromatica o ad una corrente che a seconda del modello viene tradotta nel valore glicemico corrispondente. È evidente che il campionamento non è sufficiente a rilevare tutti gli episodi che si verificano nella vita quotidiana.



Figura 1. Dispositivo pungidito per l'automonitoraggio della glicemia

Nuovi scenari nel trattamento del diabete sono stati presentati negli ultimi anni, con la creazione di dispositivi minimamente invasivi, in grado di monitorare la concentrazione del glucosio interstiziale in modo continuo (CGM, acronimo di “Continuous Glucose Monitoring”) per diversi giorni con periodo di campionamento 1-5 minuti. Il fluido interstiziale è lo spazio tra le cellule e i capillari dove avvengono scambi di sostanza, in particolare di glucosio. Ciò richiede tecniche più avanzate per lo studio delle serie temporali del glucosio, impegno compensato dai numerosi benefici ottenuti per il paziente. Analizzando il segnale in tempo reale si possono generare allarmi quando la concentrazione di glucosio supera un certo valore, prevedere ipo/iperglicemia e incorporare informazioni sull'utilizzo del glucosio, che soprattutto è fondamentale in dispositivi che mantengono la concentrazione di glucosio entro un certo intervallo, rilasciando per esempio insulina (pancreas artificiale). È necessario, quindi, che i sensori CGM diventino “intelligenti”, cioè che siano in grado di interpretare il segnale glicemico. Nel prossimo paragrafo ne illustriamo le principali caratteristiche. [5]

Il seguente grafico dimostra la principale differenza tra le tecniche SMGB e CGM: il campionamento. I puntini verdi rappresentano le misurazioni ottenute con SMGB, mentre la linea nera è stata ottenuta con CGM.

SMBG vs CGM

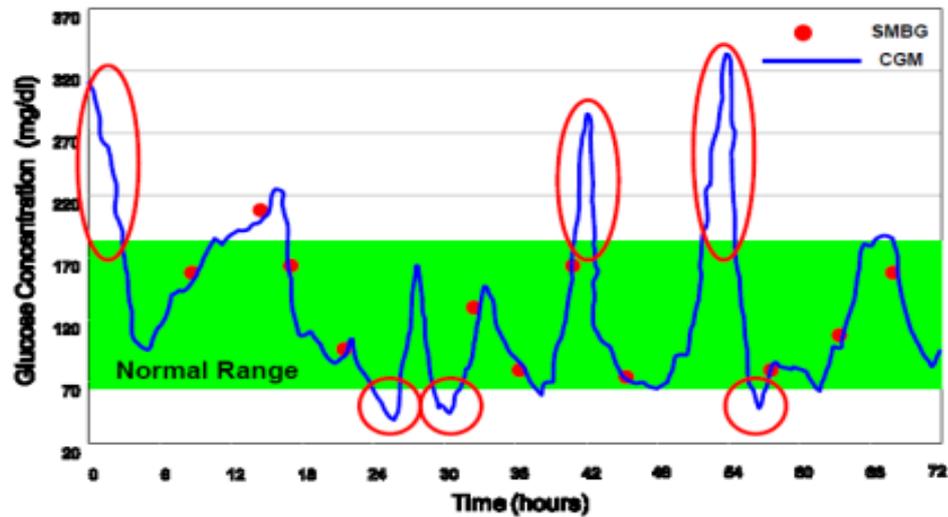


Figura 2. Confronto tra misure ottenute con SMGB e CGM. Lo spazio verde evidenziato rappresenta il range euglicemico.

Si noti che alcune escursioni, sia iperglicemiche che ipoglicemiche (cerchiate in rosso) non vengono segnalate dalle misurazioni ottenute con l'automonitoraggio della glicemia.

1.5 Sensori per monitorare la glicemia

I sensori CGM sono attualmente accettati come strumenti standard per il controllo intensivo della glicemia nei pazienti con diabete di tipo 1. Si suddividono in due categorie: biosensori mini-invasivi e non invasivi.

1.5.1 Biosensori minimamente invasivi

I sensori di glucosio basati su metodi elettrochimici rappresentano i tradizionali dispositivi per il rilevamento del CGM. Tali dispositivi sono generalmente costituiti da:

- un minuscolo sensore impiantato sottocute, generalmente nel braccio o nella zona addominale, mantenuto per un certo periodo a contatto con il liquido interstiziale;
- un trasmettitore wireless;
- un ricevitore / monitor per visualizzare il segnale acquisito.

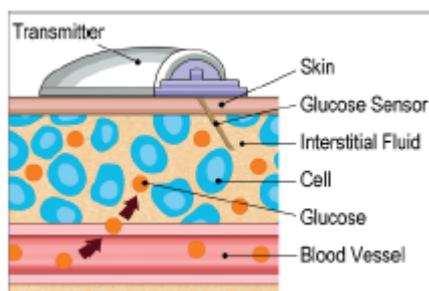


Figura3. Schema di un sensore mini-invasivo composto da un sensore di glucosio impiantato sottocute e da un trasmettitore.

La maggior parte dei sistemi minimamente-invasivi utilizzano enzimi, mentre altri sono privi di enzimi. Gli elettrodi enzimatici utilizzano enzimi per catalizzare reazioni di ossido-riduzione (redox) e il movimento degli elettroni produce una corrente o una tensione dipendente dalla concentrazione di glucosio nel sito in cui l'elettrodo è posizionato. Gli elettrodi includono micropori, microaghi, elettrodi amperometrici sottocutanei e dispositivi impiantabili per via endovenosa. Un importante progresso è stato il sensore ad ago, che generalmente consiste in un elettrodo a filo di platino-iridio (Pt-Ir) con un enzima sulla superficie e/o un rivestimento polimerico per l'impianto sottocutaneo. Importanti per questi elettrodi sono le operazioni di calibrazione del sensore, per convertire la misura di concentrazione nell'interstizio in una stima della concentrazione del glucosio nel plasma. [6]

Diversi sono gli indici proposti per la valutazione dell'accuratezza dei sensori CGM, il più popolare è l'indice di differenza relativa assoluta (MARD). Essa rappresenta la media dell'errore assoluto tra tutte le misurazioni CGM e i valori glicemici accoppiati di riferimento ottenuti con tecniche di cui vi è una riconosciuta affidabilità. Una bassa percentuale indica che le letture CGM sono vicine ai valori di riferimento, mentre una MARD in percentuale più alta indica una discrepanza più ampia con i valori di riferimento. Alla fine dell'ultimo decennio i valori di accuratezza dei sensori CGM erano significativamente peggiori rispetto ai sistemi SMGB (i quali hanno una MARD tra il 5% e il 10%). Pertanto negli ultimi 10 anni, le società hanno impiegato notevoli sforzi per migliorare i sensori e renderli più accurati.

Medtronic, nel 2011, ha lanciato il sensore Elite che ha ricevuto l'approvazione dall'FDA nel 2013, fornendo ai clienti risultati più efficaci. Infatti rispetto al dispositivo di prima generazione, la disposizione degli elettrodi è stata ridisegnata per far fronte agli effetti della risposta infiammatoria, le dimensioni del sensore sono diminuite e il sensore e il tempo di usura è stato esteso da 3 a 6 giorni. L'accuratezza di questo dispositivo è del 13,6 %. Nel 2016 la

stessa compagnia ha prodotto il sensore Guardian 3, con una MARD dell'8,7% e durata di 7 giorni.[2]



Figura 4. Sensore Elite (Medtronic) con dispositivo di inserimento.

Dexcom nel 2012 ha lanciato il nuovo sensore G4 Platinum dalla durata di 7 giorni, composto da un trasmettitore e diversi sensori che hanno permesso di aumentare la MARD al 13%; questo sensore è stato rinnovato incorporandolo con la tecnologia Share, che consente una connessione wireless sicura tra il ricevitore del paziente e un'app sullo smartphone. Nel 2015 è nato il sistema G5 Mobile CGM, che permette una comunicazione wireless diretta con uno smartphone senza la necessità di un ricevitore.



Figura 5. Sistema Dexcom G5 con trasmettitore dedicato.

Abbott nel 2011 ha messo sul mercato il sistema CGM FreeStyle Navigator 2, successore del FreeStyle Navigator, il quale dura 5 giorni, presenta un trasmettitore di dimensioni più piccole e un ricevitore con una gamma di funzioni e allarmi aggiuntivi per monitorare i livelli di glucosio. La MARD di questo dispositivo è pari al 14,5%.[7]



Figura 6. *Freestyle Navigator 2 con trasmettitore dedicato.*

1.5.2 Biosensori non invasivi

Gli attuali biosensori elettrochimici di glucosio, descritti in precedenza, sono mini-invasivi e presentano una serie di limitazioni, come il biofouling, l'incapsulamento fibroso dell'elettrodo impiantato, l'infiammazione e la perdita vascolare dell'ospite, che influiscono gravemente sulla precisione e sull'accuratezza della concentrazione di glucosio nel sangue. Un'altra barriera è la durata dei sensori, poiché la perdita di funzionalità della maggior parte degli elettrodi impiantabili si verifica entro 7 giorni e la calibrazione crea problemi di costo, tossicità e disagio. I metodi non invasivi sono sempre più diffusi a causa dell'elevata sensibilità e dell'elevata compliance del paziente, contrariamente a quelli invasivi. Sebbene siano stati implementati miglioramenti per mitigare l'imprecisione della CGM dei sensori non invasivi, non sono ancora stati approvati in via definitiva dati per la regolazione dell'insulina, poiché il suo dosaggio dipende fortemente dal livello glicemico e dalla velocità di variazione della glicemia. [6]

I sensori tipici non invasivi sono basati su differenti tecniche: estrazione ionoforetica di glucosio attraverso la pelle, risonanza plasmonica di superficie, spettroscopia Raman, spettroscopia visibile o vicino all'infrarosso (NIR), polarimetria, foto-acustiche sonde e metodi di fluorescenza. Tutti questi metodi innovativi richiedono la calibrazione dello strumento rispetto ai sensori CGM. Alcuni esempi di sensori non invasivi sono i seguenti:

- lenti a contatto, che mantengono un contatto costante con il liquido lacrimale. Il biosensore monitora il glucosio nelle lacrime, che si è visto essere correlato a quello del sangue. Le lenti a contatto devono soddisfare una serie di caratteristiche; innanzitutto non devono essere tossiche, ma devono essere morbide, costituite da un materiale con proprietà meccaniche che si mantengono inalterate nel tempo (come idrogel o polimeri plastificati) che sono più resistenti ai movimenti oculari e al battito delle palpebre.

- Lenti a contatto basate su sensori spettrali. Queste lenti a contatto sono in idrogel e incorporano una serie di particelle monodisperse, che diffrangono le onde elettromagnetiche. Gli idrogel modificati con acido fenilboronico (PBA) sono in grado di legare cis-dioli nel glucosio, portando cambiamenti volumetrici e forzando la lente ad espellere l'acqua, alterando così la percezione del colore della lente. Sono aggiunti sulle lenti anche cristalli fotonici per migliorare la sensibilità. Il comportamento di rilevamento della lente viene monitorato da uno spettrometro e i risultati hanno mostrato che la lunghezza d'onda di diffrazione è relativamente spostata con l'aumento della concentrazione di glucosio lacrimale.

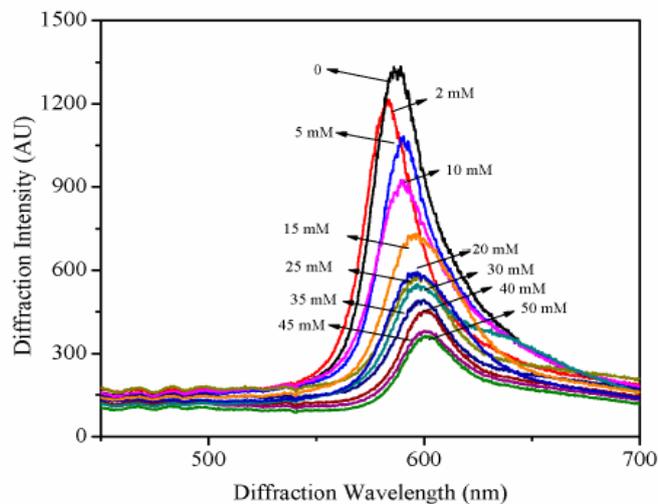


Figura 7. La lunghezza d'onda di diffrazione è shiftata con la concentrazione di glucosio nelle lacrime.

- Sensori salivari. La saliva è un ottimo fluido diagnostico, che fornisce un'alternativa alle analisi del sangue. Le concentrazioni di glucosio nella saliva variano e sono maggiori nei pazienti diabetici rispetto ai soggetti normali. Sono stati sviluppati sensori di glucosio 'paradenti', costituiti da sistemi micro-elettromeccanici (MEMS); uno tra questi è il sensore "Cavitas", che presenta un sensore di glucosio costituito da elettrodi di platino e di argento/cloruro d'argento ed incorpora l'enzima glucosio ossidasi come elemento di riconoscimento per avviare una reazione elettrochimica che produce un segnale elettrico. Un trasmettitore wireless consente la misurazione telemetrica del glucosio nella saliva.

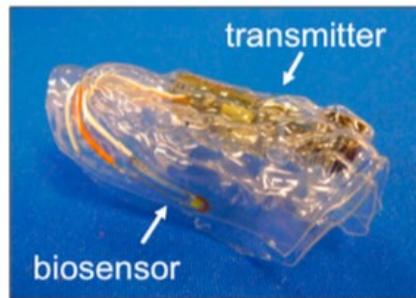


Figura 8. Immagine schematica del biosensore del paracatenti su misura per la detenzione del paziente.

- Sensori ottici. I sensori ottici utilizzano luce di frequenze variabili per rilevare il glucosio, tenendo conto delle proprietà di interazione della luce con le molecole di glucosio, dipendenti dalla concentrazione. Sulla base di questa tecnologia di trasduzione ottica, sono stati sviluppati numerosi sistemi CGM basati sull'ottica, come la fluorescenza, la spettroscopia di assorbimento a infrarossi e la spettroscopia Raman.

CAPITOLO 2: LA SUPPORT VECTOR REGRESSION

2.1 Machine learning per la predizione del glucosio nel sangue

Da quanto abbiamo trattato in precedenza risulta evidente la necessità di mantenere la concentrazione di glucosio nel sangue all' interno di un determinato range, evitando episodi di ipo/iperglicemia. La sfida consiste quindi nel prevedere i livelli futuri di glucosio nel sangue (BG) a partire dalla sua storia passata per consentire al paziente di intraprendere azioni terapeutiche. Questo interesse naturalmente è esploso nel momento in cui sono stati commercializzati i dispositivi CGM.[8]

In letteratura sono stati presentati diversi algoritmi di machine learning per prevedere il segnale glicemico; in questa tesi ci concentreremo sulla regressione lineare a vettori di supporto.

L'apprendimento automatico è lo studio scientifico di algoritmi e modelli usati dai sistemi informatici per conseguire un determinato compito senza che sia programmato in modo esplicito. Si basa quindi su diversi algoritmi per risolvere problemi di dati. Non c'è un unico algoritmo adatto a tutti i problemi, ma si devono considerare il tipo di problema da risolvere, il numero di variabili, il modello che si adatta meglio e così via.

In particolare quando si parla di apprendimento supervisionato si intende costruire un algoritmo in grado di produrre modelli e ipotesi generali utilizzando istanze fornite dall'esterno per predire il destino di istanze future. In altre parole, trovare una funzione che associa un input in ingresso ad un output sulla base di coppie input-output preesistenti, detti campioni di addestramento.



Figura 9. Schema di funzionamento di un algoritmo di machine learning. In ingresso vengono forniti dei dati di input, i quali vengono elaborati da un algoritmo di machine learning e viene prodotto in uscita un output.

Uno dei più famosi modelli di apprendimento supervisionato sono le support vector machines (SVM). Questi modelli sono saldamente radicati nel quadro della teoria dell'apprendimento statistico, o teoria VC, che è stata sviluppata negli ultimi tre decenni da Vapnik e Chervonenkis. SVM tratta fondamentalmente due tipologie di problemi: regressione e classificazione. Quindi esistono due categorie principali di macchine a vettori di supporto: la classificazione a vettori di supporto (SVC) e la regressione a vettori di supporto (SVR). Il primo esegue la classificazione, prevedendo etichette categoriali discrete e quindi come output ha un insieme finito, il secondo esegue la regressione prevedendo in output variabili ordinate continue. [9]

2.2 Concetti generali sulle support vector machines

Per capire l'essenza degli algoritmi SVM è necessario capire il significato dei seguenti termini nella figura seguente.

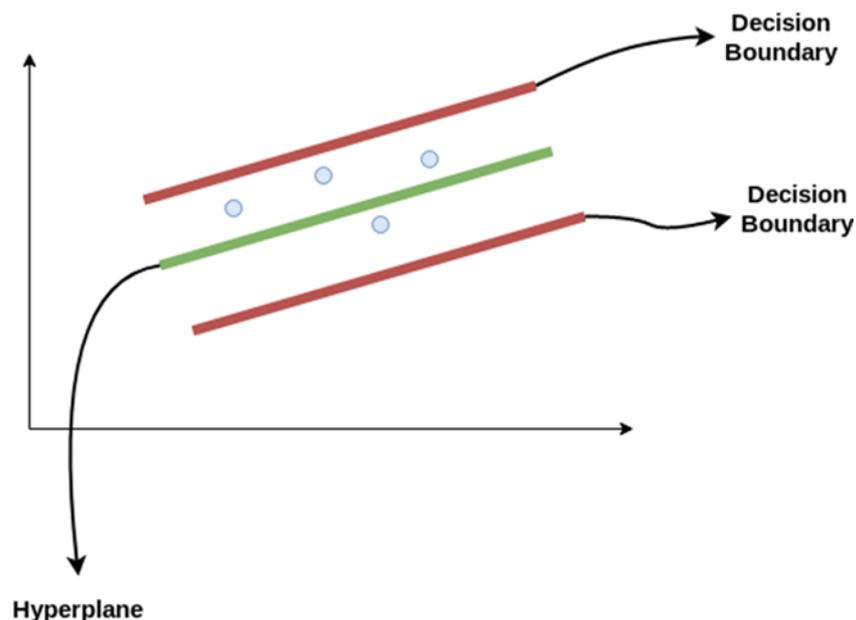


Figura 10. Caratteristiche comuni degli algoritmi a vettori di supporto: iperpiano, vettori di supporto e confine decisionale.

- *Iperpiano* (linea centrale in figura): è un sottospazio la cui dimensione è minore di una unità rispetto allo spazio ambiente. Separa lo spazio in due parti.

Esempi di iperpiano:

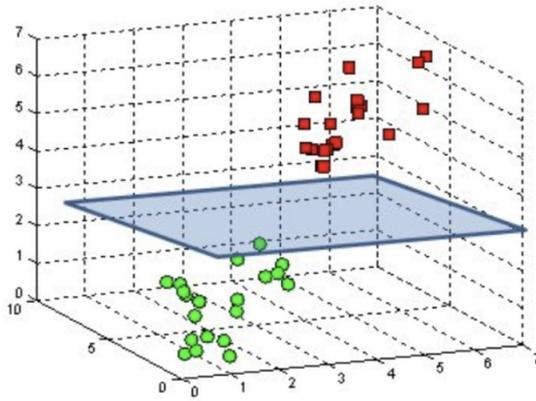


Figura 11: Spazio ambiente tridimensionale, per cui l'iperpiano ha dimensione bidimensionale.

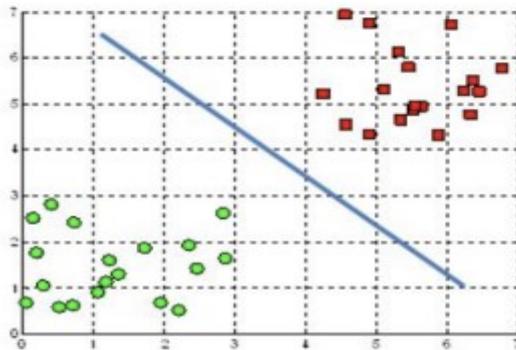


Figura 12: Spazio ambiente bidimensionale, per cui l'iperpiano è monodimensionale.

In quest'ultima figura l'iperpiano è la linea che divide due classi (in un problema di classificazione binaria). Nella regressione a vettori di supporto, è la linea che sarà utilizzata per predire l'output continuo.

- *vettori di supporto* (punti in figura): sono i dati che sono più vicini alle linee di confine, dipendono dal set di dati analizzato, influenzano la posizione e l'orientamento dell'iperpiano.
- *limite decisionale* (linee rosse in figura): sono le due linee sopra e sotto l'iperpiano, creano il margine. La distanza di ciascuna linea dall'iperpiano è solitamente indicata con ϵ .

2.3 Scopo delle support vector machines

Come detto prima, gli algoritmi supervisionati di machine learning si dividono in algoritmi di classificazione e regressione.

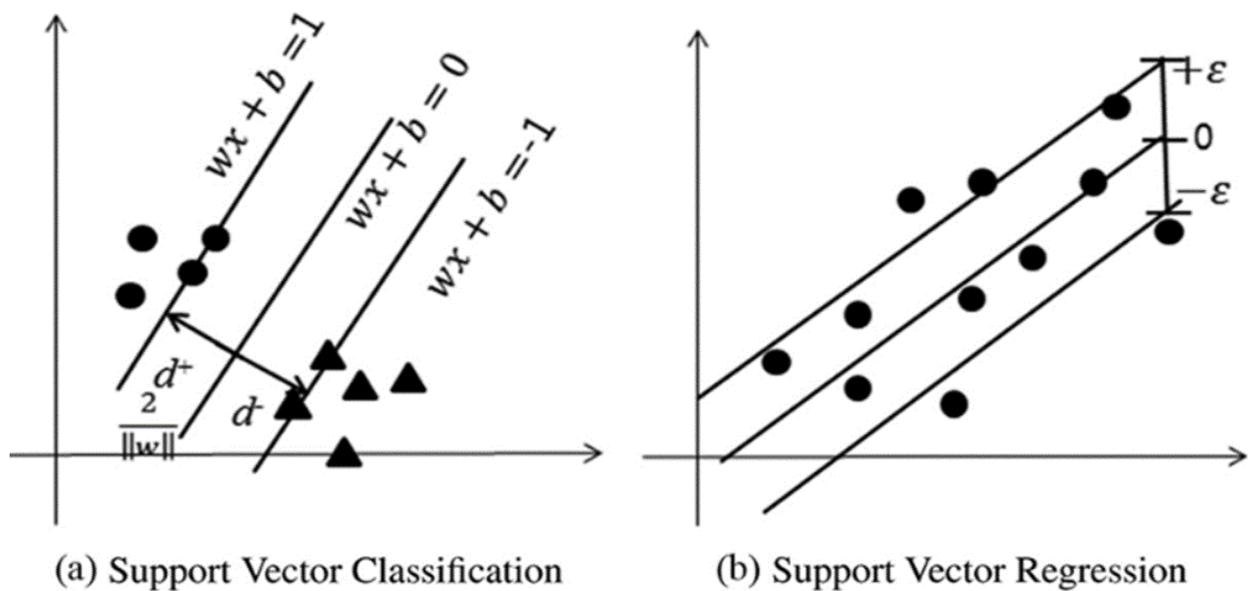


Figura 13. A destra, viene fornito un esempio del problema di classificazione, a sinistra un esempio del concetto di regressione.

In figura A abbiamo un esempio di classificazione. Ci sono due tipologie di dati e quindi avremo due classi (classe triangolo e classe cerchio). L'obiettivo è trovare l'iperpiano che restituisce il massimo margine per fare la classificazione. I punti sul margine sono i vettori di supporto e devono essere il più distante possibile fra loro

In figura B ci sono dati aventi le stesse caratteristiche (tutti cerchi). L'obiettivo è trovare l'iperpiano in modo che la maggior parte dei dati sia all'interno del margine, quest'ultimo è definito dal valore di ϵ , che rappresenta la tolleranza.

2.4 Cenni su support vector classifier (SVC)

SVR è un'estensione dell'algoritmo di classificazione, per cui vediamo l'idea di base di quest'ultimo.

2.4.1 SVC dal punto di vista geometrico

L'apprendimento di classificazione mira a trovare l'equazione di una superficie che separa due o più classi in maniera ottimale nello spazio delle caratteristiche considerato. Consideriamo ora la seguente figura e supponiamo di avere un insieme di punti che appartengono a due classi, e di voler separarle in modo da assegnare correttamente i punti futuri ad una o ad un'altra classe. SVM si preoccupa di trovare l'iperpiano che separa le due classi con il margine più alto possibile. A volte risulta difficile separare le due classi e si potrebbero avere valori anomali che cadono all'interno del margine. Nel grafico seguente abbiamo dei punti blu e verdi, per separarli vediamo che:

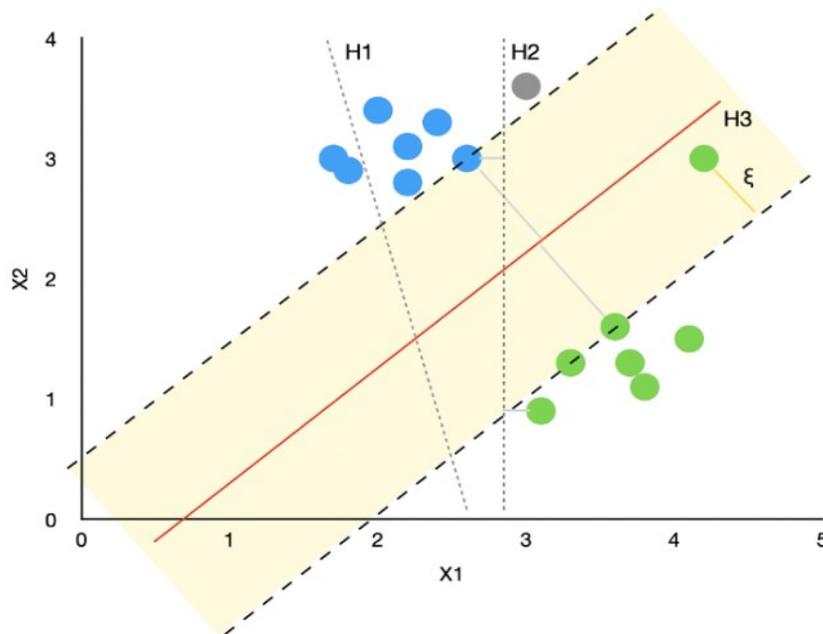


Figura 14. Separare le due classi di punti con algoritmo SVM.

- H1 non riesce a separare le classi correttamente;
- H2 separa le classi correttamente, ma la distanza tra la retta H2 e i punti più vicini è abbastanza piccola. Sarà più probabile classificare nuovi punti futuri in modo errato.
- H3 divide le due classi con il margine più alto possibile, rendendo il modello molto più robusto. Ciò permette che nuovi punti futuri vengano con buona probabilità classificati in maniera corretta. Si osservi che un punto si trova all'interno del margine [10]

2.4.2 Principali proprietà di SVC

- SVM è una tecnica sparsa. Quando vengono presi i parametri per il modello SVM è necessario che i dati di allenamento siano disponibili, cioè archiviati tutti in memoria. Tuttavia, dopo aver identificato i parametri, SVM dipende solo da un insieme di queste istanze di addestramento, chiamati vettori di supporto. Essi definiscono i margini degli iperpiani.
- Kernel di SVM. SVM usa il trucco del kernel per mappare i dati in uno spazio di dimensione superiore prima di risolvere il compito di apprendimento automatico.
- SVM è un separatore di margine massimo. L'iperpiano trovato deve essere posizionato alla massima distanza dalle diverse classi, ciò è essenziale perché i dati di addestramento rappresentano solo un campione di popolazione, mentre la previsione deve essere eseguita su istanze che non abbiamo ancora che possono avere una distribuzione leggermente diversa da quella del sottoinsieme di addestramento.[11]

2.4.3 Controllo della dimensione di Vapnik-Chervonenkis (VC)

Si considerino funzioni che valgono soltanto +1 o -1, dette funzioni indicatrici:

$$h: X \rightarrow \{-1, +1\}$$

Queste funzioni dividono gli elementi di X in due classi dette dicotomie, come si può vedere nella seguente figura

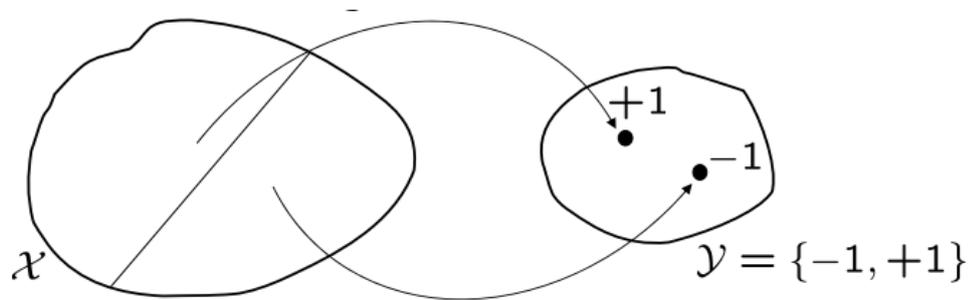


Figura 15: Funzioni indicatrici.

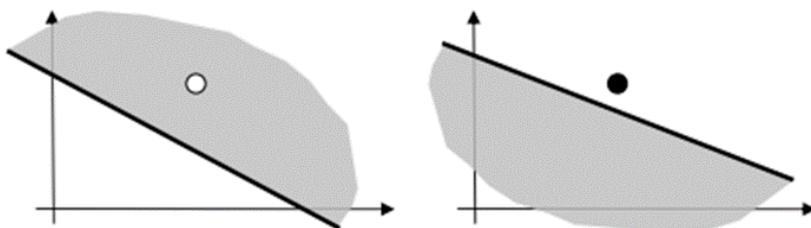
Sia Z un sottoinsieme di X , Z è frammentato (shattered) dall'insieme H se H realizza tutte le possibili dicotomie di Z .

La dimensione delle funzioni indicatrici H ed è detta dimensione-VC (dal nome dei suoi ideatori Vapnik-Chervonenkis) ed è uguale a quella del sottoinsieme di X che ha cardinalità maggiore.

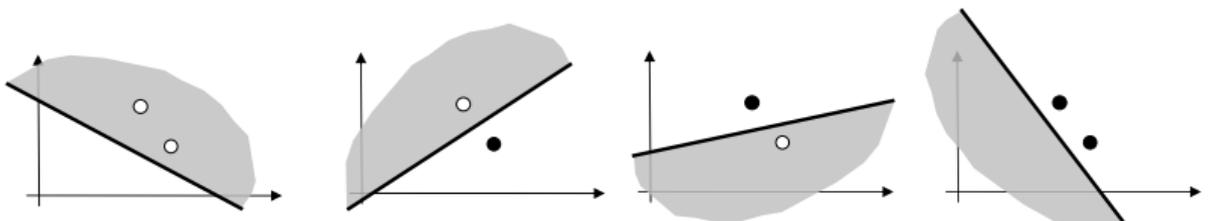
$$VC(H) = \max \{|Z| : Z \text{ è frammentato da } H\}, Z \subseteq X$$

Esempi:

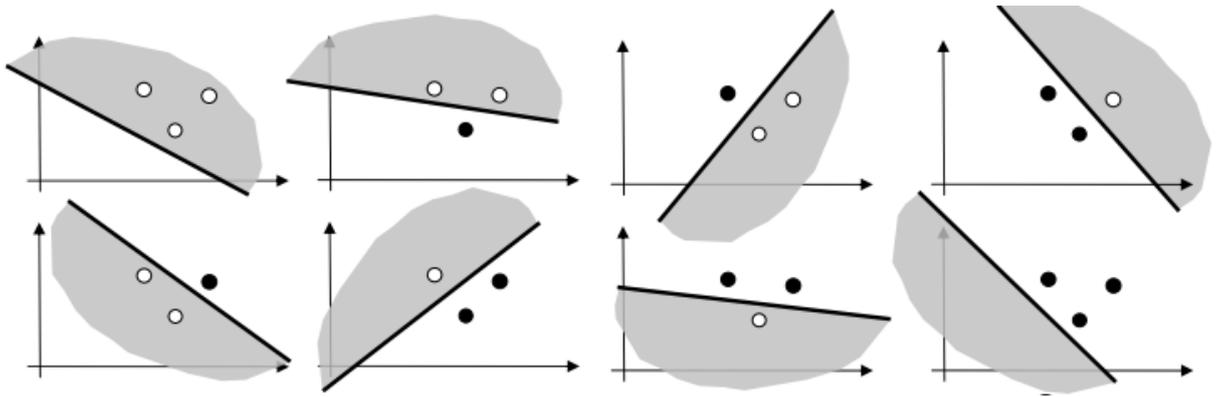
-Con un solo punto si hanno $2^1 = 2$ dicotomie:



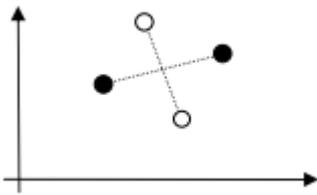
-Con due punti si hanno $2^2 = 4$ dicotomie:



-Con due punti si hanno $2^3 = 8$ dicotomie:



-nessuna disposizione di quattro punti può essere divisa da piani, per cui $VC(H)=3$:



Generalmente la dimensioni VC di un insieme di iperpiani separatori in dimensione n è di $n+1$.

2.4.2 Hard Margin e Soft margin

L'idea principale di un hard margin ('margine rigido') è di massimizzare il margine sotto il vincolo di non compiere errori di classificazione. Nella figura notiamo che non ci sono punti all'interno del margine.

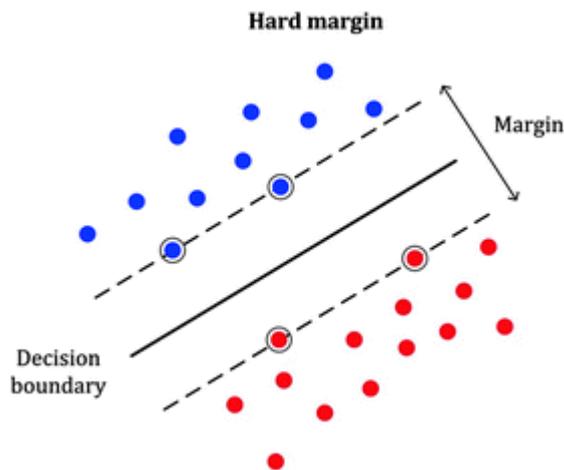


Figura 16. Esempio di hard margin.

Quando i dati non sono completamente separati, sono introdotte le variabili slack per consentire errori di classificazione. In questo caso SVM non ha margine rigido che classificherebbe tutti i dati in modo impeccabile, ma ha un margine detto 'soft'. Il modello classificherebbe in modo corretto la maggior parte dei punti, e allo stesso tempo classificherebbe erroneamente alcuni punti in prossimità del confine di separazione. Nella figura seguente si noti che un punto rosso si trova

nella regione dell'iperpiano sbagliata e verrà classificato erroneamente come blu; lo stesso accade per il punto rosso che verrà predetto blu

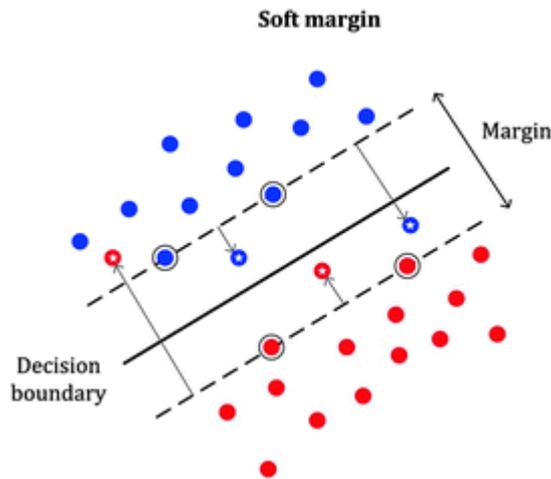


Figura 17. Esempio di soft margin.

2.5 Support vector regression (SVR)

Come la classificazione, la regressione a vettori di supporto è una tecnica sparsa, caratterizzata dall'uso del kernel e dal controllo di CV e del numero di vettori di supporto.

2.5.1 Principi di funzionamento

Il problema della regressione è trovare una funzione che approssimi la mappatura da un dominio di input a numeri reali sulla base di un campione di addestramento. Supponiamo che l'iperpiano sia una linea retta con equazione: $Y=wx+b$ e che linee di confine siano a una distanza di $+\epsilon$ e $-\epsilon$ da essa. Qualsiasi iperpiano che soddisfi l'algoritmo SVR dovrebbe soddisfare:

$$-\epsilon \leq Y-wx-b \leq \epsilon$$

L'obiettivo principale è decidere un confine decisionale a una distanza ' ϵ ' dall'iperpiano in modo tale che i punti dati più vicini all'iperpiano o ai vettori di supporto si trovino all'interno della linea nera di confine. Il confine decisionale è il margine di tolleranza e prenderemo solo i punti che si trovano all'interno di questo confine, cioè quelli che hanno il minor tasso di errore. L'obiettivo principale è quindi minimizzare l'errore individuando l'iperpiano che massimizza il margine, tenendo in mente che parte dell'errore è tollerato.

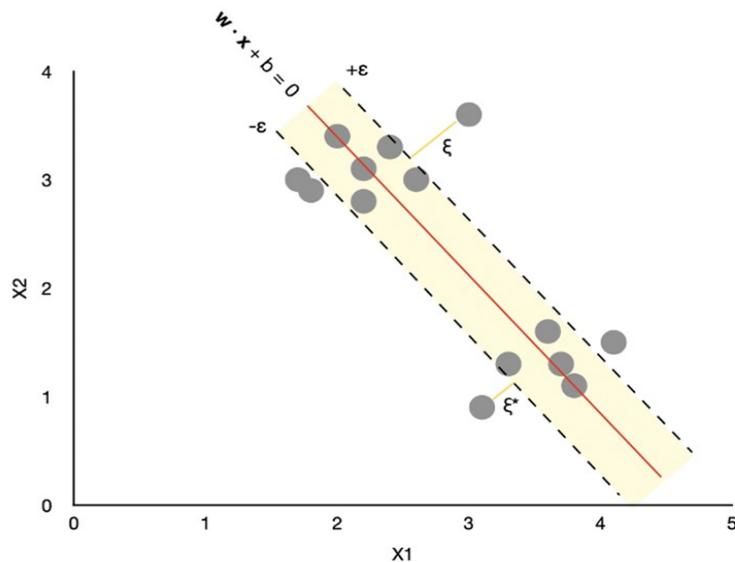


Figura 18. Iperpiano e confine decisionale.

A differenza di SVM, si introduce quindi una regione ϵ -insensibile attorno alla funzione, chiamata ϵ -tubo. Si tratta quindi di trovare il tubo più stretto che meglio si avvicina alla funzione a valore continuo stimata centrato attorno a quest'ultima, riducendo al minimo l'errore di previsione ovvero la distanza tra l'uscita prevista e quella desiderata (problema di ottimizzazione). Infine le variabili slack misurano la distanza dei punti al di fuori del tubo, e per considerarli nel seguito si introdurrà il parametro di regolarizzazione C .

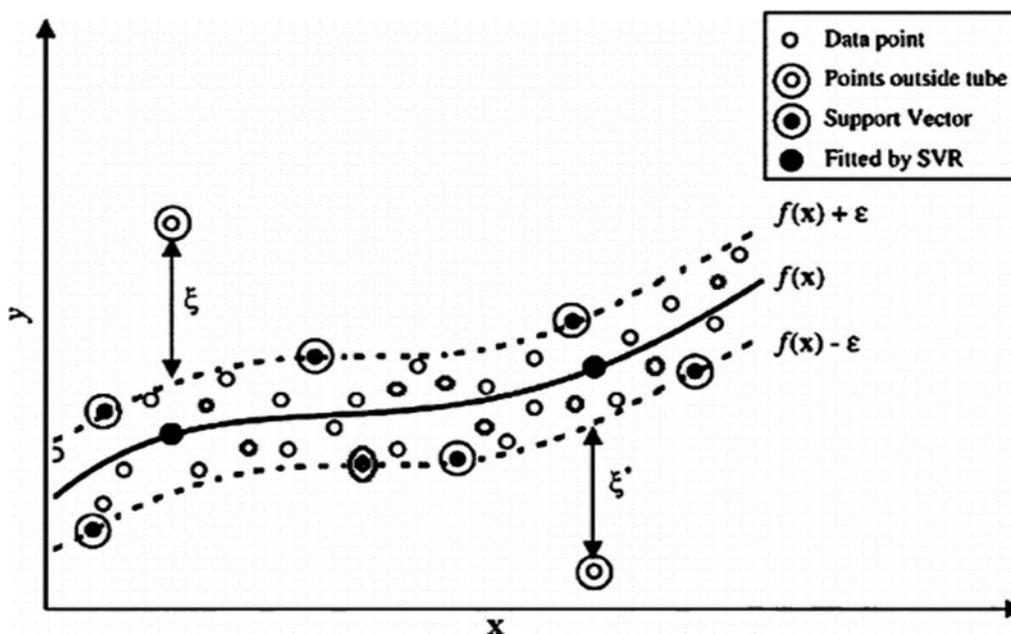


Figura 19: Introduzione regione ϵ -insensibile al problema di regressione.

2.5.2 Significato del parametro w dell'iperpiano di separazione

Per calcolare il margine di un iperpiano di separazione, è necessario calcolare la distanza tra l'iperpiano e il punto più vicino all'iperpiano.

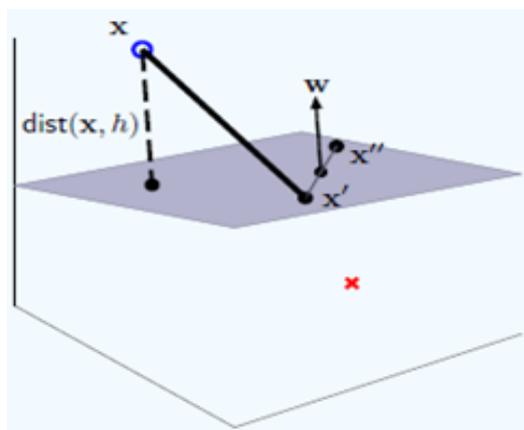


Figura 20. Rappresentazione grafica della distanza di un punto da un iperpiano.

Come inizio calcoliamo la distanza $\text{dist}(x, h)$ tra un punto arbitrario x dall'iperpiano di separazione $h = (b, w)$. Questa distanza è la lunghezza del segmento perpendicolare che parte da x e finisce in h . Chiamiamo x' un punto qualsiasi che giace nell'iperpiano, quindi che soddisfa l'equazione dell'iperpiano $w^T x' + b = 0$. Con u indichiamo il vettore unitario che è normale all'iperpiano. Vale la relazione:

$$\text{dist}(x, h) = |u^T(x - x')|$$

Ossia la distanza di x da h è la proiezione del vettore $(x - x')$ su u . Sosteniamo ora che w è perpendicolare all'iperpiano per cui poniamo $u = w / \|w\|$. Perciò la relazione prima diventa:

$$\text{dist}(x, h) = |u^T(x - x')| = \frac{|w^T x - w^T x'|}{\|w\|} = \frac{|w^T x + b|}{\|w\|}$$

Dove abbiamo usato $w^T = -b$. Consideriamo tutti i punti x_1, \dots, x_n , quelli più vicini all'iperpiano sono:

$$\min_{n=1, \dots, N} \text{dist}(x_n, h) = \frac{1}{\|w\|}$$

Per ogni iperpiano di separazione, il margine è $1/\|w\|$. [12]. Ci occuperemo ora di trovare l'iperpiano il cui margine contiene il maggior numero di dati.

2.5.3 Modello matematico per la regressione lineare

Il modello prodotto dalla classificazione a vettori di supporto dipende solo da un sottoinsieme dei dati di addestramento, perché non ci si preoccupa dei punti di addestramento che si trovano oltre il margine. I dati di addestramento sono stati presi come $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset \mathbb{N} \times \mathbb{R}$, dove \mathbb{N} indica lo spazio dei modelli di ingresso, ad esempio \mathbb{R}^d . L'obiettivo è trovare una funzione $f(x)$ che abbia la massima deviazione ε dal valore reale target y_i per tutti i dati di addestramento e allo stesso tempo sia la più piatta possibile. Iniziamo considerando funzioni lineari del tipo:

$$y = f(x) = \langle w, x \rangle + b = \sum_{j=1}^M w_j x_j + b, \quad y, b \in \mathbb{R}, x, w \in \mathbb{R}^M$$

In caso di dati multidimensionali la notazione diventa:

$$f(x) = \begin{bmatrix} w \\ b \end{bmatrix}^T \begin{bmatrix} x \\ 1 \end{bmatrix} = w^T x + b \quad x, w \in \mathbb{R}^{M+1}$$

Per funzione piatta significa che si cerca una w piccola. Un modo per garantire questo è minimizzare la norma di w . Minimizzare la norma di w significa allo stesso tempo trovare il massimo margine di separazione. Possiamo riscriverlo come un problema di ottimizzazione convessa:

$$\min_w \frac{1}{2} \|w\|^2$$

Condizioni al contorno

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases}$$

Per convenienze matematiche si è utilizzato $\frac{1}{2}$ e si è elevato al quadrato. Si noti che le condizioni al contorno costituiscono un hard margin.

Questo problema di ottimizzazione è analogo a quello per la classificazione perché nessun tipo di violazione oltre il margine è ammessa. Per cui a volte non c'è nessuna soluzione per questo problema e ciò comporta il ricorso ad un soft margin. [13]

Per visualizzare come w influenzi la planarità della funzione, si consideri il seguente esempio:

$$f(x, w) = \sum_{i=1}^M w_i x^i, \quad x \in \mathbb{R}, w \in \mathbb{R}^M$$

dove M è l'ordine del polinomio usato per approssimare una funzione. Più il vettore w è grande, più w_i sono diversi da 0, risultando in soluzioni di ordine superiore. La linea

orizzontale rappresenta la soluzione polinomiale di ordine 0 ed è molto distante dall'output desiderato, quindi ha un errore grande. La funzione lineare produce approssimazioni migliori solo per una parte dei dati, ma distante rispetto ai dati di addestramento. La soluzione di sesto ordine, anche se più complessa, rappresenta il miglior compromesso tra la planarità della funzione e l'errore di previsione. W ci consente quindi di regolare la planarità della soluzione. [11]

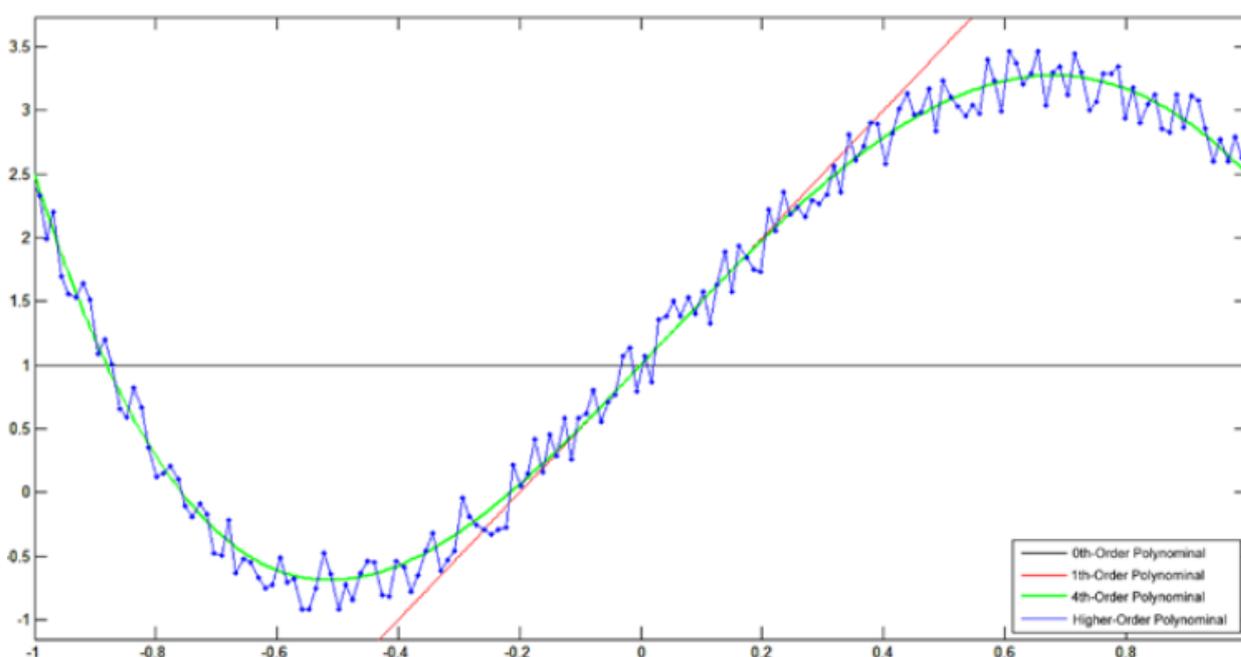


Figura 21. Grado delle soluzioni.

2.5.4 Funzione di perdita

SVR adotta una funzione di perdita ϵ -insensibile, fino ad una certa soglia l'errore è nullo, mentre tutte le deviazioni dall'output desiderato, sia alte che basse, ricevono una penalità. Il valore di ϵ determina la larghezza del tubo. Se ϵ è piccolo il tubo è spostato verso l'interno, pertanto più punti sono attorno al confine e ci sono più vettori di supporto. Allo stesso modo, aumentando ϵ ci saranno meno punti attorno al confine. Le funzioni di perdita più utilizzate sono la funzione lineare, quadratica e di Huber. La funzione di Huber è più uniforme di quella lineare e quadratica, ma penalizza le deviazioni dall'output desiderato con una penalità maggiore a mano a mano che aumenta l'errore. Tutte e tre sono simmetriche e possono essere adottate per limitare la sottostima o la sopravvalutazione.

$$L_\varepsilon(y, f(x, w)) = \begin{cases} 0 & |y - f(x, w)| \leq \varepsilon; \\ |y - f(x, w)| - \varepsilon & \text{otherwise,} \end{cases}$$

$$L_\varepsilon(y, f(x, w)) = \begin{cases} 0 & |y - f(x, w)| \leq \varepsilon; \\ (|y - f(x, w)| - \varepsilon)^2 & \text{otherwise,} \end{cases}$$

$$L(y, f(x, w)) = \begin{cases} c|y - f(x, w)| - \frac{c^2}{2} & |y - f(x, w)| > c \\ \frac{1}{2}|y - f(x, w)|^2 & |y - f(x, w)| \leq c \end{cases}$$

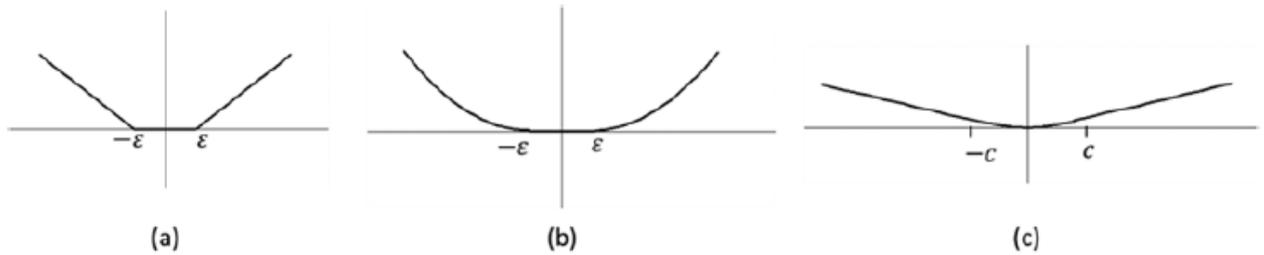


Figura 22. Funzione di perdita lineare(a), quadratica(b), di Huber(c).

In questa trattazione useremo la funzione di perdita lineare.

2.5.5 Introduzione delle variabili slack

Fin qui abbiamo considerato che il margine sia rigido, il che a volte comporta che non ci sia nessuna soluzione al problema di ottimizzazione. Analogamente alla classificazione, anche qui possiamo introdurre il soft margin per ammettere errori al di fuori del margine. Per far fronte ad essi si introducono allora le variabili slack ξ_i, ξ_i^* , che determinano quanti punti possono essere tollerati all'esterno del tubo illustrato in figura.[12][13]. Si assuma che f effettivamente esista e che approssimi tutte le coppie (x_i, y_i) con precisione la formulazione diventa:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i + \xi_i^*$$

Con le seguenti condizioni:

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Dalle condizioni al contorno si nota che ora vengono accettati tutti i punti all'interno del margine più alcuni all'esterno.

Andremo a penalizzare le variabili slack aggiunte nella sommatoria regolando l'iperparametro C . All'aumentare di C , aumenta anche la tolleranza per i punti al di fuori di ϵ . Quando C si avvicina a 0, la tolleranza si avvicina a 0 e l'equazione crolla in quella semplificata.

Se non si sono ottenuti i risultati richiesti nemmeno passando dall'hard margin al soft margin, allora si può ricorrere all'uso del kernel.

2.6 Kernel

A volte è necessario spostarsi da uno spazio con certe caratteristiche ad un altro di diverse caratteristiche. Nel precedente paragrafo abbiamo considerato che la funzione $f(x)$ fosse lineare. Se la funzione non è lineare si può applicare il trucco del kernel, che rende la regressione a vettori di supporto una tecnica molto flessibile. Il kernel è una funzione che prende il problema non lineare originale e lo trasforma in uno lineare, che viene poi gestito dall' algoritmo in uno spazio a dimensione superiore (chiamato "feature space").

Ciò si può fare preelaborando i dati di addestramento x_i tramite mappatura:

$$\phi: X \rightarrow \mathfrak{S}$$

e poi applicare l'algoritmo SVR.

Considerando la seguente figura che rappresenta un problema di classificazione, si può notare che le due classi nello spazio di input non sono linearmente separabili ma attraverso la funzione di mappatura i dati vengono mappati in uno spazio in cui sono linearmente separabili.

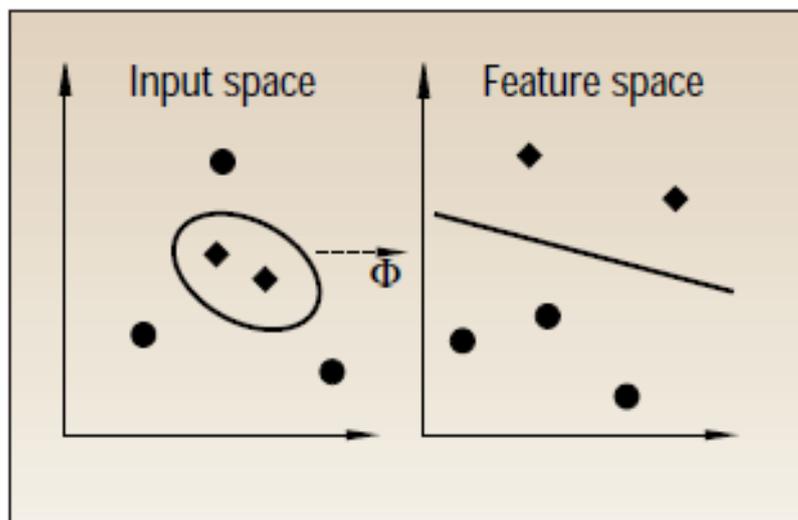


Figura 23. Mappare un set di dati di addestramento non lineare in uno spazio a dimensione superiore e costruire un iperpiano con il massimo margine.

Analogamente, in caso di un problema di regressione, in cui i dati di input non sono convenienti perché porterebbero alla costruzione di una funzione non lineare, si può utilizzare una funzione di mappatura in modo che la funzione stimata finale che meglio approssima il valore target sia lineare.

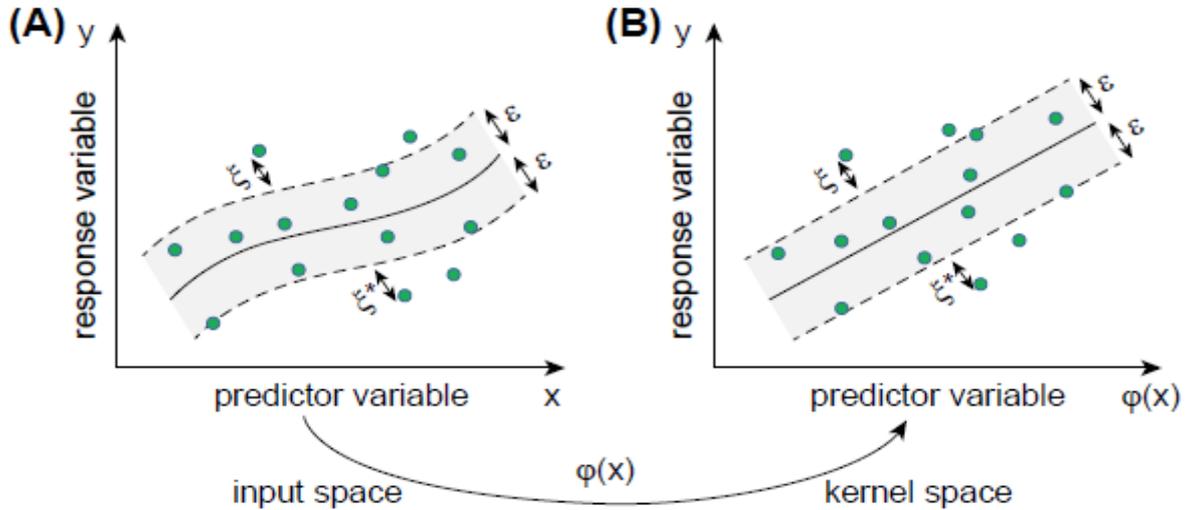


Figura 24. Mappare un set di dati di addestramento non lineare in uno spazio a dimensione superiore in cui il margine dell'iperpiano contenga il maggior numero di dati.

Questa tecnica risulta utile in algoritmi che utilizzano i campioni di addestramento attraverso prodotti scalari $x_i \cdot x$. In questo caso non è necessario conoscere le singole funzioni $\phi(x_i)$ e $\phi(x)$, ma è sufficiente il prodotto scalare di $\phi(x_i) \cdot \phi(x)$. Il kernel è dato dal prodotto:

$$k(x_i, x) = \phi(x_i) \cdot \phi(x)$$

I kernel più usati sono:

- Lineare: $K(x_i, x) = x_i \cdot x$
- Polinomiale: $K(x, x_i) = (x \cdot x_i)^d$
- Gaussian Radial Basis function: $K(x, x_i) = \exp(-|x - x_i|^2 / (2\sigma^2))$
- Sigmoid: $K(x, x_i) = \tanh(kx \cdot x_i - \delta)$

2.7 Rischio funzionale

Si considerino i dati di addestramento del seguente insieme:

$$\mathbf{X} := \{(x_1, y_1), \dots, (x_\ell, y_\ell)\} \subset \mathcal{X} \times \mathbb{R}$$

Si assuma che le coppie di dati di addestramento siano indipendenti fra loro e uniformemente distribuiti, con una determinata distribuzione di probabilità $P(x,y)$. Tra tutte le funzioni possibili è importante scegliere f in modo che minimizzi l'errore teorico, detto anche errore atteso, dato da:

$$R[f] = \int c(x, y, f(x)) dP(x, y)$$

$c(x,y,f(x))$ è detta funzione di costo, determina quanto vengono penalizzati gli errori di stima sulla base dei dati empirici \mathbf{X} . Il rischio atteso indica quanto il valore predetto y_i per un punto x sia distante dal valore reale. $P(x,y)$ è incognita, per cui non è possibile calcolarla esplicitamente. Una soluzione è approssimare il rischio empirico con il funzionale di rischio empirico, tenendo conto che i campioni classificati in precedenza (quelli appartenenti all'insieme \mathbf{X}) rappresentano un campione di $P(x,y)$. Si ottiene allora:

$$R_{\text{emp}}[f] := \frac{1}{\ell} \sum_{i=1}^{\ell} c(x_i, y_i, f(x_i)).$$

Ora non compare più $P(x,y)$, quindi è facile calcolarlo. Per avere l'errore minimo teorico è necessario minimizzare l'errore empirico, ma allo stesso tempo, rendere minima la dimensione VC. L' algoritmo SVR è in grado di risolvere questo problema. Viene infatti aggiunto un termine di controllo $\|w\|^2$ per regolarizzare il rischio funzionale:

$$R_{\text{reg}}[f] := R_{\text{emp}}[f] + \frac{\lambda}{2} \|w\|^2$$

in cui $\lambda > 0$ è detta costante di regolarizzazione. [14]

CAPITOLO 3: APPLICAZIONE DI SVR PER LA PREDIZIONE DEL SEGNALE DI GLUCOSIO IN CONTINUA

3.1 Caso di studio: predizione multivariata della concentrazione di glucosio sottocutaneo in pazienti affetti da diabete di tipo 1 basata su Support Vector Regression

Tramite questo studio mostriamo come viene applicata la regressione a vettori di supporto. I risultati ottenuti dall'algoritmo sono fortemente influenzati dal tipo e dal numero di variabili di input che vengono tenute in considerazione e per questo verranno distinti 6 casi. Maggiore è il numero di input esogeni (assunzione di carboidrati, somministrazione di insulina, sforzi fisici ...), più lungo sarà l'orizzonte di previsione. Le principali variabili di input sono: il profilo glicemico, la concentrazione plasmatica di insulina, la comparsa di glucosio nella circolazione sistemica derivante dai pasti e il dispendio energetico durante attività fisiche.

3.1.1 Dati

I dati sono stati raccolti da un progetto di ricerca dell'Unione Europea denominato METABO. I pazienti diabetici sono 27 e tutti seguono una terapia insulinica. Il periodo di osservazione dello studio va dai 5 a 22 giorni. In tabella seguente sono riportate le caratteristiche dei pazienti, con il numero medio di eventi ipo/iperglicemici e la loro durata media. Si è considerato un episodio ipoglicemico quando due valori di glucosio consecutivi sono inferiori a 60 mg/dl; quando due valori consecutivi sono al di sopra di 180 mg/dl si parla invece di iperglicemia. Il tempo di campionamento utilizzato coincide con quello del sensore CGM utilizzato in questo studio (Guardian Real-Time CGM di Medtronic) ed è di 5 min. Le informazioni relative al dispendio energetico sono state acquisite da una fascia indossabile che monitora segnali fisiologici derivanti da più sensori.

I pazienti si sono poi preoccupati di scrivere in un diario giornaliero le informazioni relative ai cibi assunti, le dosi di insulina e l'istante temporale di iniezione.

I 27 pazienti sono stati suddivisi in tre gruppi perché alcuni non avevano la fascia indossabile oppure altri non avevano correttamente scritto per ogni giorno i pasti o l'insulina assunta. Il gruppo A comprende 15 pazienti di cui abbiamo tutte le informazioni; al gruppo B appartengono cinque pazienti per i quali non si hanno i dati relativi all'energia persa durante le attività fisiche; infine il gruppo C comprende i restanti sette pazienti dei quali si ha soltanto il segnale CGM.

Patient Baseline Characteristics		Descriptive Statistics of the Glucose Dataset	
Gender		Average	0.40
No. Female	12	Hypoglycemic	(0-2)
No. Male	15	Events Per Day	
Age (years)		Average Duration	43.10
Mean \pm SD	43.5 \pm 13.4	of Hypoglycemic	(0-88.33)
Range	19-72	Events (min)	
BMI (kg/m ²)		Average	1.64
Mean \pm SD	25 \pm 3.70	Hyperglycemic	(0-3.36)
Range	18.75–35.80	Events Per Day	
HbA1c (%)		Average Duration	129.30
Mean \pm SD	7.07 \pm 1.11	of Hyperglycemic	(0-283.75)
Range	5.20 – 8.50	Events (min)	

Tabella 25. Dati. I valori nelle parentesi indicano il minimo e il massimo valore dei pazienti.

Vengono poi esaminati 6 casi di predizione del glucosio sottocutaneo a seconda delle variabili di input utilizzate:

- Caso 1: come input vengono utilizzate le misurazioni in continua delle concentrazioni di glucosio passate (gl);
- Caso 2: alle misurazioni in continua del glucosio vengono aggiunte la concentrazione di insulina nel plasma (I_p) e il rate di comparsa del glucosio derivante nel plasma dai pasti (R_a);
- Caso 3: gli input del caso 2 sono arricchiti con l'informazione della quantità totale di glucosio esogeno presente nella circolazione sistemica ($S R_a$);
- Caso 4: a tutte le variabili di input precedenti è stata aggiunto il tempo a cui sono state rilevate (h);
- Caso 5: al caso 3 è aggiunta l'energia persa durante l'attività fisica quotidiana (SEE);
- Caso 6: anche il caso 4 è stato arricchito con l'informazione dell'energia persa nell'attività fisica.

I casi sono confrontati fra loro e il risultato predittivo viene valutato dal punto di vista dell'accuratezza.

3.1.2 Algoritmo

Per predire i livelli di glucosio per ogni paziente viene utilizzata la tecnica SVR combinata ai modelli compartimentali del sistema di regolazione glucosio-insulina. Supponiamo che la concentrazione futura di glucosio sottocutaneo al tempo $t+1$, con t che indica il tempo corrente, sia espressa dalla seguente funzione $f: \mathbb{R}^d \rightarrow \mathbb{R}$

$$f(x) = f(x_1, \dots, x_n)$$

dove le x_i indicano le variabili di input esplicitate prima (gl, I_p , Ra, SRa, h, SEE); il valore di n è compreso tra 1 e 6 dato che vengono utilizzate al massimo 6 variabili di input. Rispetto al momento t in cui viene fatta la previsione viene considerata la storia passata delle variabili di input. Ogni variabile di input x_i , eccetto h, è descritta da un vettore di lunghezza finita contenente valori successivi nell'intervallo di tempo:

$$[t_{x_i} - n_{x_i} \Delta t_{x_i}, t_{x_i}]$$

dove

$$x_i = [x_i(t_{x_i} - n_{x_i} \Delta t_{x_i}), \dots, x_i(t_{x_i} - \Delta t_{x_i}), x_i(t_{x_i})]$$

t_{x_i} indica il limite superiore, Δt_{x_i} è il periodo di campionamento e n_{x_i} è la lunghezza della finestra temporale. Dunque, il numero totale di input (numero di caratteristiche) del modello è:

$$d = \sum_{i=1}^n (n_{x_i} + 1).$$

In base alla tecnica SVR, la funzione di predizione è data dalla seguente formula lineare:

$$f(x) = w^T \phi(x) + b$$

$\phi(x)$ indica una trasformazione fissata nello spazio delle caratteristiche, per mappare i dati di allenamento x^j , con j che denota la dimensione dei dati di addestramento ($j=1, \dots, N$) in uno spazio a diversa dimensione dove la relazione tra x^j e gli output y_i è lineare. Viene aggiunta la funzione di perdita ϵ -insensibile, in cui l'errore aumenta linearmente a mano a mano che ci si allontana nella regione ϵ -insensibile. Quando gli errori superano il confine $\pm\epsilon$ si introducono le variabili di slack ξ_j, ξ_j^* e il problema di ottimizzazione diventa:

$$\text{Minimize } C \sum_{j=1}^N (\xi_j + \xi_j^*) + \frac{1}{2} \|w\|^2$$

Con i vincoli

$$\begin{cases} y_j \leq f(x^j) + \varepsilon + \xi_j \\ y_j \geq f(x^j) - \varepsilon - \xi_j^* \\ \xi_j, \xi_j^* \geq 0 \end{cases}$$

Risolvendo il problema di ottimizzazione si trova che la funzione di predizione è:

$$f(x) = \sum_{j=1}^N (a_j - a_j^*) k(x, x^j) + b$$

a_j, a_j^* sono i moltiplicatori di Lagrange.

3.1.3 Valutazione del modello

Il metodo viene eseguito per ciascun paziente utilizzando il set di dati acquisiti durante il periodo di osservazione. Sono stati utilizzati quattro orizzonti di previsione: 15, 30, 60 e 120 min. Si osservi che il caso 1 può essere eseguito su tutti i pazienti, i casi 2, 3 e 4 possono essere eseguiti solo sui pazienti dei gruppi A e B, i casi 5 e 6 possono essere applicati solo al gruppo A.

Per valutare le prestazioni del modello viene utilizzata una procedura di convalida incrociata dieci volte. Come kernel viene utilizzato quello gaussiano, mentre gli iperparametri C , ε e il parametro del kernel γ sono ottimizzati con l'algoritmo differenziale evolutivo (DE). I parametri vengono ottimizzati in modo che il modello si adatti il più possibile ai dati di addestramento. La funzione di fitness per l'algoritmo DE è definita dall'RMSE, cioè la radice del quadrato medio del processo di convalida incrociato 10 volte

$$\text{RMSE}_{\text{tenfold}} = \frac{1}{10} \sum_{k=1}^{10} \sqrt{\frac{1}{N_k} \sum_{j=1}^{N_k} (y_j - f(x^j))^2}$$

Dove N_k rappresenta la dimensione di kth-fold, y_j è il valore di glucosio associato all'input x^j e $f(x^j)$ è il valore di glucosio ottenuto con l'algoritmo SVR.

Viene utilizzato anche il coefficiente r_{tenfold} :

$$r_{\text{tenfold}} = \frac{1}{10} \sum_{k=1}^{10} r_k$$

Sia $\text{RMSE}_{\text{tenfold}}$ che r_{tenfold} sono utilizzati per valutare l'accuratezza delle previsioni.

3.1.4 Ottimizzazione di parametri

Nella seguente tabella sono riportati il valore medio e la deviazione standard dei parametri C, ϵ e γ del modello SVR riguardanti il gruppo A.

Case #	Prediction Horizon											
	15 min			30 min			60 min			120 min		
	C	ϵ	γ									
Case 1	557.82 (255.96)	0.68 (0.24)	6.07 (2.45)	593.33 (326.57)	0.46 (0.33)	6.48 (1.54)	516.96 (274.67)	0.45 (0.29)	7.08 (2.02)	622.39 (346.46)	0.57 (0.33)	7.69 (0.70)
Case 2	801.27 (166.08)	0.33 (0.24)	4.48 (1.58)	986.58 (66.05)	0.47 (0.34)	7.50 (0.70)	953.05 (112.33)	0.69 (0.26)	7.91 (0.22)	983.41 (65.73)	0.39 (0.37)	7.93 (0.14)
Case 3	863.94 (186.85)	0.51 (0.32)	3.71 (1.68)	875.72 (197.68)	0.35 (0.21)	6.48 (1.28)	947.15 (114.77)	0.47 (0.35)	7.83 (0.38)	938.27 (124.89)	0.52 (0.35)	7.89 (0.24)
Case 4	861.07 (182.47)	0.42 (0.28)	3.26 (1.14)	839.81 (194.82)	0.50 (0.37)	6.22 (1.58)	880.29 (125.73)	0.51 (0.29)	7.28 (0.81)	954.92 (132.61)	0.46 (0.32)	7.87 (0.50)
Case 5	707.07 (217.32)	0.49 (0.25)	1.68 (0.63)	791.40 (157.37)	0.37 (0.25)	2.74 (0.91)	873.90 (169.41)	0.40 (0.31)	3.75 (1.73)	867.76 (211.22)	0.33 (0.30)	4.73 (2.48)
Case 6	801.49 (176.96)	0.51 (0.32)	1.53 (0.66)	875.63 (130.78)	0.32 (0.24)	2.58 (0.94)	791.80 (208.67)	0.47 (0.36)	3.34 (1.65)	850.48 (179.60)	0.48 (0.29)	4.74 (2.48)

Figura 26. Ottimizzazione dei parametri del modello predittivo SVR (gruppo A).

Si noti che i parametri sono diversi per ciascun caso di input e per orizzonte di previsione. Tuttavia, si possono fare alcune considerazioni. Il valore medio di C nei casi 2 e 6 è più alto rispetto al caso 1, sia per un orizzonte di previsione breve, sia per uno un po' più lungo. Ciò significa che si ottengono previsioni più fluide a scapito della complessità del modello. Il parametro ϵ nella maggior parte dei dati è compreso tra 0.32 e 0.57, a parte nei casi 1 e 2 in cui è circa 0.7 per predizioni di 15 e 60 min. Si osserva poi che ϵ è minore nei casi 5 e 6, che indica una migliore capacità di generalizzare il modello. Tuttavia ϵ diventa più grande più l'orizzonte di previsione cresce, per cui l' 'algoritmo è meno uniforme per previsioni lunghe. Per i gruppi 1 e 4 non ci sono sostanziali differenze.

3.1.5 Risultati dell'applicazione

Nella tabella seguente sono mostrati il valore medio di $RMSE_{tenfold}$ e $t_{tenfold}$ e la loro deviazione standard.

Case#	Prediction Horizon							
	15 min		30 min		60 min		120 min	
	RMSE	r	RMSE	r	RMSE	r	RMSE	r
Group A								
Case 1	9.05 (2.24)	0.98 (0.01)	15.29 (2.76)	0.95 (0.01)	24.19 (3.78)	0.87 (0.05)	33.04 (6.49)	0.72 (0.12)
Case 2	7.31 (1.74)	0.99 (0.00)	9.15 (1.88)	0.98 (0.01)	12.24 (2.89)	0.97 (0.02)	15.31 (4.27)	0.94 (0.04)
Case 3	6.55 (1.74)	0.99 (0.00)	8.24 (2.25)	0.99 (0.01)	10.55 (2.85)	0.97 (0.02)	13.43 (4.20)	0.95 (0.03)
Case 4	6.10 (1.83)	0.99 (0.00)	7.16 (2.01)	0.99 (0.01)	8.56 (2.23)	0.98 (0.01)	10.33 (2.99)	0.97 (0.02)
Case 5	5.35 (1.63)	0.99 (0.00)	6.38 (1.57)	0.99 (0.00)	7.82 (2.01)	0.99 (0.01)	8.57 (1.75)	0.98 (0.01)
Case 6	5.21 (1.58)	0.99 (0.00)	6.03 (1.67)	0.99 (0.00)	7.14 (1.84)	0.99 (0.01)	7.62 (1.81)	0.99 (0.01)
Group B								
Case 1	9.56 (0.67)	0.98 (0.00)	16.60 (1.94)	0.95 (0.01)	27.51 (5.07)	0.86 (0.04)	37.56 (7.34)	0.72 (0.07)
Case 2	7.45 (1.21)	0.99 (0.00)	9.83 (1.39)	0.98 (0.01)	13.65 (2.48)	0.96 (0.02)	16.26 (3.33)	0.95 (0.03)
Case 3	6.92 (1.23)	0.99 (0.00)	8.66 (1.65)	0.99 (0.01)	11.90 (1.84)	0.97 (0.01)	14.28 (2.50)	0.96 (0.01)
Case 4	6.03 (0.94)	0.99 (0.00)	6.96 (1.04)	0.99 (0.00)	7.77 (1.57)	0.99 (0.00)	10.00 (2.01)	0.98 (0.01)
Group C								
Case 1	10.44 (2.26)	0.97 (0.02)	17.30 (2.58)	0.92 (0.05)	26.73 (3.53)	0.81 (0.06)	37.45 (6.82)	0.57 (0.10)

Figura 27. RMSE e r per valutare la precisione della predizione.

Nel gruppo A, per il caso 1 si osserva che per bassi orizzonti di previsione del glucosio, l'errore è basso mentre la correlazione è alta. L'accuratezza, come si può già intuire, diminuisce poi per le previsioni a 60 e 120 min. Nel caso 2, per le previsioni a 15 e 30 min, l'RMSE medio è 10 volte inferiore a 10mg/dl, invece per orizzonti di previsione più ampi, è simile a quello del caso 1. Nel caso 3, avendo una variabile di input aggiuntiva rispetto ai casi precedenti, l'RMSE diminuisce anche per le previsioni di 60 e 120 min. Il caso 4 dimostra che il tempo è una variabile significativa per migliorare la predizione. I risultati ottimali si sono avuti nei casi 5 e 6 con l'aggiunta della variabile SEE, in cui l'errore di previsione è inferiore a 8 mg/dl per ogni orizzonte di previsione.

Di seguito sono riportati i grafici del segnale glicemico relativi ai casi 1, 3 e 6 per orizzonti di previsione di 30 e 60 minuti visti nell'arco di una giornata. La curva blu rappresenta la concentrazione reale di glucosio nel sangue, quella rossa è la concentrazione predetta.

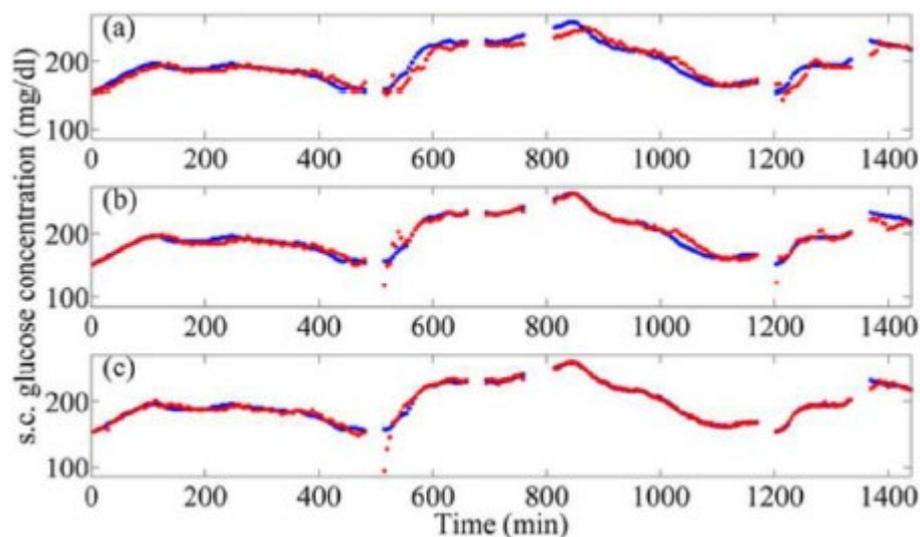


Figura 28. Concentrazione di glucosio reale e prevista dall' algoritmo SVR nel caso 1(a), caso 2(b) e caso 3(c) per una previsione di 30 min.

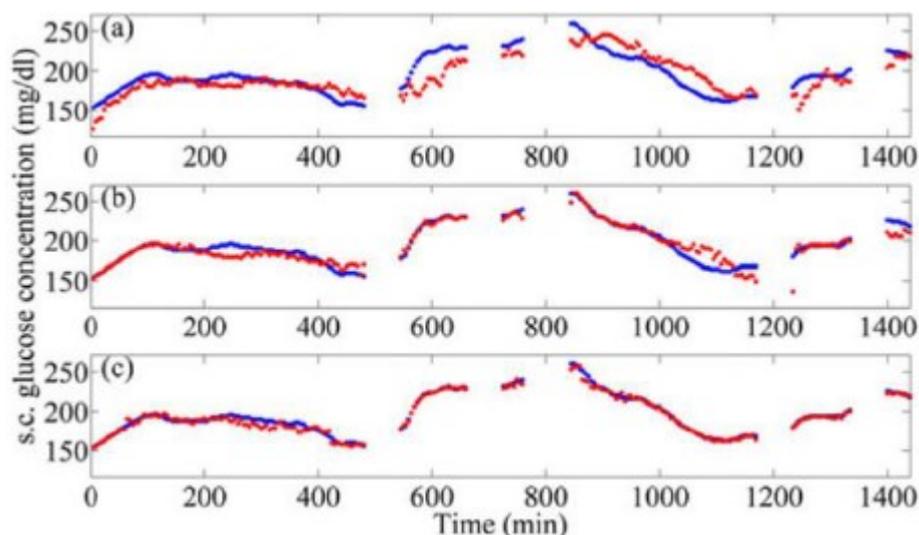


Figura 29. Concentrazione di glucosio reale e prevista dall' algoritmo SVR nel caso 1(a), caso 2(b) e caso 3(c) per una previsione di 60 min.

Dai grafici si può vedere come le curve si sovrappongono maggiormente fra loro passando dal primo al terzo caso.

Nella tabella in figura 30 sono contenuti i risultati relativi al gruppo A per i casi 4 e 6 riguardante le previsioni corrette, le previsioni errate benigne (cioè che probabilmente non portano a esito negativo) e le previsioni errate.

	Prediction Horizon											
	15 min			30 min			60 min			120 min		
	AR	BE	ER	AR	BE	ER	AR	BE	ER	AR	BE	ER
Hypoglycemia												
Case 4	96.76	2.00	1.24	94.56	2.56	2.88	92.16	2.74	5.09	87.63	4.50	7.88
Case 6	96.75	2.37	0.89	94.05	2.93	3.02	90.59	3.91	5.50	90.05	6.32	3.63
Euglycemia												
Case 4	95.84	3.79	0.37	95.51	4.20	0.28	94.35	5.33	0.33	93.03	6.65	0.32
Case 6	96.56	3.16	0.29	95.80	3.91	0.29	94.93	4.81	0.25	93.57	6.24	0.20
Hyperglycemia												
Case 4	89.68	6.09	4.23	88.20	7.76	4.04	86.45	9.53	4.02	84.85	11.03	4.12
Case 6	90.00	5.82	4.19	89.28	6.57	4.15	87.45	8.75	3.81	83.29	12.98	3.73

Figura 30. AR: previsioni corrette, BE: previsioni errate benigne, ER: previsioni errate.

Più del 90% delle previsioni per qualsiasi orizzonte vengono classificate come corrette oppure come errate, ma benigne. Si noti poi che le previsioni a breve termine per gli eventi ipoglicemici sono più accurate rispetto a quelle degli eventi iperglicemici e allo stesso tempo, l'accuratezza diminuisce per l'ipoglicemia più l'orizzonte di previsione diventa ampio. Se in tabella ci fossero stati anche i dati relativi ai i casi 1, 2 e 3, le previsioni errate sarebbero state in maggior numero.

Anche per i gruppi B e C il modello predittivo funziona, però è necessaria una maggiore attenzione a causa dei parametri liberi (assunzione dei pasti, attività fisica, ecc...). [15]

CONCLUSIONI

Nello studio precedente è stato mostrato come diversi parametri di input endogeni e esogeni combinati assieme influiscono sulla capacità di predire il segnale di glucosio. Si è visto essere decisiva in SVR la scelta degli iperparametri C ed ϵ e dei parametri del kernel, i quali però aumentano il costo computazionale nella risoluzione del problema di ottimizzazione. I risultati dello studio hanno dimostrato che utilizzando come variabile di input il solo segnale glicemico (caso 1) si ottengono risultati ottimi per orizzonti di previsione brevi (massimo 30 minuti). Sfruttando tutte le variabili di input come nel caso 6, invece, l'MRSE medio è 10 volte superiore al caso 1 per orizzonti di previsione brevi. Come si poteva intuire, in tutti i casi, aumentando l'orizzonte di previsione, si è assistito a una diminuzione della precisione. Per previsioni più lunghe, il caso 1 potrebbe portare anche a numerosi potenziali errori, mentre negli altri casi si ottengono previsioni clinicamente accettabili. Dal confronto del caso 4 e 6 si può anche vedere che l'informazione sul dispendio di energia durante l'attività fisica, è importante ma non indispensabile per ottenere una buona predizione.

Questo studio presenta comunque dei limiti, perché ci sono dei fattori che non sono stati presi in considerazione. Per esempio nelle variabili di input abbiamo considerato la concentrazione di insulina e la dose iniettata senza tener conto delle condizioni di iniezione (sito di iniezione, temperatura corporea), oppure per quanto riguarda il pasto, non si è considerata la sua composizione (fibre, grassi...). Inoltre si deve tenere in considerazione che l'algoritmo è stato applicato usando dati ricavati da una determinata popolazione quindi non descrive la totale variabilità dei pazienti. Si potrebbe pensare nel futuro di sfruttare anche input esogeni riguardanti dettagli sullo stile di vita di un paziente (turni di lavoro notturni, viaggi frequenti, stato psicofisico, età...) in modo da rendere l'algoritmo più efficace e personalizzato.

La predizione degli eventi di ipo/iperglicemia fornita dall'algoritmo SVR analizzato in questa tesi è quindi estremamente utile, poiché può offrire al paziente la capacità di allerta. Se si verifica un evento critico il paziente può, ad esempio, ingerire una certa quantità di zucchero per prevenire ipoglicemie, oppure effettuare iniezioni di insulina in caso di iperglicemie. L'algoritmo fornisce una stima accurata del valore futuro della concentrazione di glucosio nel sangue, sfruttando l'informazione disponibile dai campioni passati di glicemia rilevati dai sensori di monitoraggio. Ciò comporta che il metodo SVR sia già integrato in numerosi sistemi di monitoraggio e gestione del diabete come parte dei sistemi di supporto alle decisioni.

BIBLIOGRAFIA

- [1] **David L. Nelson, Michael M. Cox.** I principi di biochimica di Lehninger, sesta edizione, Zanichelli.
- [2] **Giacomo Cappon, Giada Acciaroli, Martina Vettoretti, Andrea Facchinetti, Giovanni Sparacino.** Wearable Continuous Glucose Monitoring Sensors: A Revolution in Diabetes Treatment. *Electronics* 2017, 6(3), 65; <https://doi.org/10.3390/electronics6030065>
- [3] **Denise R. Ferrier.** Le basi della biochimica, seconda edizione italiana condotta sulla sesta edizione americana
- [4] **Grazia Ferrai1 , Renata Paleari2 per il Gruppo di Studio Intersocietario SIBioC-SIMeL Diabete Mellito.** Razionale e frequenza dell'automonitoraggio glicemico nei pazienti, diabetici https://sibioc.it/wp-content/uploads/2021/05/710_2007-automonitoraggio-RP-BC-1.pdf
- [5] **Giovanni Sparacino, Andrea Facchinetti and Claudio Cobelli.** “Smart” Continuous Glucose Monitoring Sensors: On-Line Signal Processing Issues, *Sensors* **2010**, *10*, 6751-6772; doi:10.3390/s100706751
- [6] **Cheng Chen, Xue-Ling Zhao, Zhan-Hong Li, Zhi-Gang Zhu 1, Shao-Hong Qian and Andrew J. Flewitt.** Current and Emerging Technology for Continuous Glucose Monitoring, DOI: 10.3390/s17010182
- [7] Abbott Diabetes Care Inc. Immagine di Abbott FreeStyle Navigator II. Disponibile online: <http://www.diabete.co.uk/cgm/abbottfreestyle-navigator-ii.html>
- [8] **Matteo Gadaleta, Andrea Facchinetti, Enrico Grisan, Michele Rossi.** Prediction of Adverse Glycemic Events from Continuous Glucose Monitoring Signal DOI:10.1109/JBHI.2018.2823763
- [9] **Batta Mahesh.** Machine Learning Algorithms -A Review. DOI:10.21275/ART20203995
- [10] **Saul Dobilas,** SVM Classifier and RBF Kernel — How to Make Better Models in Python, Jan 17,2021,Towards Data Science
- [11] **Mariette Awad and Rahul Khanna.** Efficient Learning Machines, Support Vector Regression, chapter 4, Apress open

- [12] **Abu-Mostafa, Magdon-Ismail**, Support Vector Machines, e-chapter 8 , Gennaio 2015, <https://edisciplinas.usp.br/pluginfile.php/5078086/course/section/5978681/chapSVM.pdf>
- [13] **Debasish Basak, Srimanta Pal and Dipak Chandra Patranabis**, Support Vector Regression
- [14] **Alex J. Smola and Bernhard Scholkopf**, A tutorial on support vector regression, 2004 Kluwer Academic Publishers. Manufactured in The Netherlands
- [15] **E I Georga, V C Protopappas, D Ardigo, M Marina, I Zavaroni, D Polyzos, D I Fotiadis**, Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression, DOI: 10.1109/TITB.2012.2219876