Analysis of transient states in EMG signals with applications to robotic prostheses

Martina Giacobbe

Università di Bologna

23 Marzo 2018

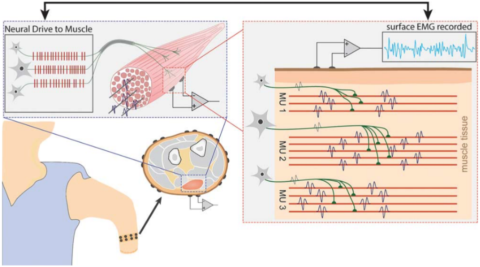
Relatrice:Prof.ssa Giovanna Citti
Correlatori:Prof. Davide Barbieri
Dott. Emre Baspinar
Ing. Emanuele Gruppioni

Obbiettivo della tesi

Oggetto di studio: Segnale EMG, ovvero la manifestazione elettrica di un'attivazione neuromuscolare associata ad una contrazione muscolare.

⇒ In particolare siamo interessati allo *stato transiente* del segnale.

Obbiettivo: Ottimizzare i tempi di risposta del controllo protesico per amputati transradiali.



Principal component analysis

Dati T campioni di un vettore random $x = (x_1, \dots, x_n)$ di componenti correlate tale che $E[x] = 0$, definiamo

$$y_1 = \sum_{k=1}^{N} w_{k1} x_k = w_1^T x$$

è detta prima componente principale di x se $\mathrm{Var}(y_1)$ è massima.
Obbiettivo: Massimizzare

$$J(w_1) = var(y_1) = w_1 C_x w_1$$

sotto il vincolo $\|w_1\| = 1$.
Si ottiene $w_1 = e_1 \quad \Rightarrow \quad y_1 = e_1^T x$ prima componente principale.

Generalizzando: se $y_m = m$-esima componente principale $\Rightarrow$ la condizione di scorrelazione con le componenti principali precedenti porta ad un problema di ottimizzazione analogo al precedente in $span\{e_1, \dots, e_{m-1}\}^\perp$

$$w_m = e_m \quad \Rightarrow \quad y_m = e_m^T x$$

Principal component analysis

Osservazione

Massimizzare $\text{Var}(y_1) \quad \Leftrightarrow \quad$ *Minimizzare il MSE tra* x *e* $\bar{x}$

ove $\bar{x} = \sum_{i=1}^{M} (w_i^T x) w_i$.

Infatti

$$J_{MSE} = E[\|x - \bar{x}\|^2] = \sum_{i=M+1}^{N} w_i C_x w_i \quad \Rightarrow \quad J_{MSE} = \sum_{i=M+1}^{N} \lambda_i$$

affinché J_{MSE} sia minimizzato $\lambda_{M+1}, \dots, \lambda_N$ devono essere gli autovalori più piccoli. $\Rightarrow x$ viene proiettato nello spazio generato da $e_1, \dots, e_m$. Dunque:

$$w_i = e_i \quad \text{e} \quad Var(y_i) = E[e_i^T x x^T e_i] = d_i.$$

Support vector machine: caso linearmente separabile

$\{(x_i, d_i)\}$ training set, ove x_i campione e $d_i = \pm 1$ etichetta.

Obbiettivo: Trovare w_o vettore dei pesi e b_o bias affinché l'iperpiano

$$w_o^T x_i + b_o = 0$$

massimizzi il margine di separazione ρ, ove ρ rappresenta la distanza tra l'iperpiano e il punto del dataset più vicino. La separabilità lineare si riassume con la condizione

$$d_i(w_o^T x_i + b_o) \geq 1 \qquad \text{per } i = 1, \ldots, N \tag{1}$$

I punti del dataset che soddisfano (1) con l'uguaglianza sono detti *support vectors*.

Support vector machine: caso linearmente separabile

Definisco $g(x) = w_o^T x + b_o$ e riscrivo $x = x_p + r\frac{w_o}{\|w_o\|}$, ove x_p è la proiezione ortogonale di x sull'iperpiano. Usando $g(x_p) = 0$ si ottiene

$$g(x) = w_o^T x + b_o = w_o^T \left(x_p + r\frac{w_o}{\|w_o\|} \right) + b_o = r\|w_o\|$$

$$\Rightarrow r = \frac{g(x)}{\|w_o\|}$$

Chiamo $x^{(s)}$ un *support vector* con etichetta $d^{(s)}$, vale
$g(x^{(s)}) = w_o^T x^{(s)} + b_o = \pm 1, \text{ per } d^{(s)} = \pm 1.$ Dunque

$$r = \frac{g(x^{(s)})}{\|w_o\|} = f(n) = \begin{cases} \frac{1}{\|w_o\|} & \text{if } d^{(s)} = +1 \\ -\frac{1}{\|w_o\|} & \text{if } d^{(s)} = -1. \end{cases}$$

Infine $\rho = 2r = \frac{2}{\|w_o\|}$.

Osservazione

Massimizzare $\rho \quad \Leftrightarrow \quad$ *Minimizzare* $\|w_o\|$

Support vector machine: caso linearmente separabile

Teorema (Problema primario)

Dato il training set $\{(x_i, d_i)\}$, *il problema primario si occupa di trovare i valori di ottimo di* w *e* b *tali che soddisfino il vincolo* $d_i(w^T x_i + b) \geq 1$ *e* w *minimizzi* $\phi(w) = \frac{1}{2} w^T w$.

Teorema (Problema duale)

Dato il training set $\{(x_i, d_i)\}$, *il problema duale si occupa di trovare i valori di ottimo dei moltiplicatori di Lagrange* $\{\alpha_i\}_{i=1}^N$ *tali che soddisfino i vincoli* $\sum_{i=1}^N \alpha_i d_i = 0$ *e* $\alpha_i \geq 0, \ \forall i = 1, \dots, N$ *e massimizzino* $Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j$

$\Rightarrow w_o = \sum_{i=1}^{N_s} \alpha_{o,i} d_i x_i$

$\Rightarrow b_o = 1 - w_o^T x^{(s)} = 1 - \sum_{i=1}^{N_s} \alpha_{o,i} d_i x_i^T x^{(s)} \quad \text{con } d^{(s)} = 1$

Support vector machine: caso non linearmente separabile

Definizione

Il margine di separazione ρ è detto essere soft se almeno un punto del dataset (x_i, d_i) non soddisfa

$$d_i(w^T x_i + b) \geq 1 \qquad \text{for } i = 1, \dots, N.$$

Si introducono delle variabili di slack $\{\xi_i\}_{i=1}^N$, tali che

$$d_i(w^T x_i + b) \geq 1 - \xi_i \qquad i = 1, \dots, N. \tag{2}$$

$0 < \xi_i \leq 1 \quad \Rightarrow \quad$ classificazione corretta
$\xi_i > 1 \quad \Rightarrow \quad$ classificazione errata

Teorema (Problema primario)

Dato il training set $\{(x_i, d_i)\}$, il problema primario si occupa di trovare i valori di ottimo w_o e b_o tali che soddisfino il vincolo in (2) e $\xi_i \geq 0$, $\forall\ i$, inoltre le variabili di slack $\{\xi_i\}_{i=1}^N$ e w minimizzino la funzione

$$\Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

SVM come kernel machine

Sia x vettore nello spazio di input di dimensione m_0, sia poi $\{\phi_j(x)\}_{j=1}^{\infty}$ un setting di funzioni non lineari. Attuando tale trasformazione, l'equazione dell'iperpiano diventa

$$\sum_{j=1}^{\infty} w_j \phi_j(x) = 0 \quad \Leftrightarrow \quad w^T \phi(x) = 0,$$

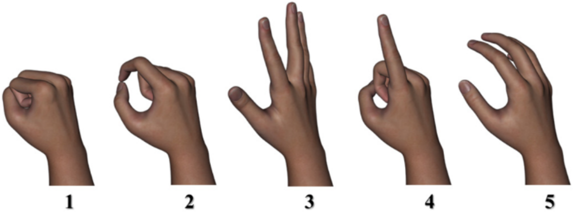
ove $\{w_j\}_{j=1}^{\infty}$ setting di pesi.
Adattando la soluzione del problema primario al nostro caso si ha $w = \sum_{i=1}^{N_s} \alpha_i d_i \Phi(x_i)$

$$\Rightarrow \sum_{i=1}^{N_s} \alpha_i d_i \Phi^T(x_i)\Phi(x) = 0.$$

Si definisce

$$k(x, x_i) = \Phi^T(x_i)\Phi(x) = \sum_{j=1}^{\infty} \phi_j(x_i)\phi_j(x) \quad \text{per } i = 1, \dots, N_s$$

Registrazione del segnale



Registrazione del segnale: $P = 5, \quad R = 10, \quad T = 4s, \quad Fr = 500Hz$
Ripetizione: matrice $T \cdot Fr \times 6 \quad \rightarrow \quad 2000 \times 6$
Campione: vettore 1×6
Dataset: matrice $P \cdot R \cdot Fr \cdot T \times 6$

Eliminazione dei picchi

Definizione

Un valore discreto del segnale all'istante t è detto un picco se la differenza tra il valore discreto in t e il valore discreto in $t-1$ è maggiore di 10^4.

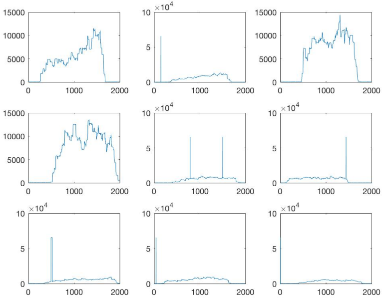


Figura: Posa 1, sensore 1, rappresentazione al variare delle ripetizioni.

Preprocessamento del segnale

Calcoliamo le soglie di riposo di ogni sensore

$$TH = [79, 149, 103, 818, 190, 363].$$

Pose 1,2,3,4: Enumero le ripetizioni e salvo l'indice temporale di troncamento in un vettore D $40 \times 1 \Rightarrow$ estraggo $max(D)$.
Ridimensiono una ripetizione $Datum_i$ 2000×6 con troncamento all'istante $D(i)$
Se $D(i) = \max(D) \Rightarrow Datum_i = Datum_i(max(D) : end - 1, :)$
Altrimenti $\Rightarrow Datum_i = Datum_i(D(i) : end - max(D) + D(i) - 1, :)$
Posa 5: Una ripetizione $Datum_{rest}$ 2000×6 viene ridimensionata come

$$Datum_{rest} = Datum_{rest}(1 : end - max(D), :)$$

Preprocessamento del segnale

Le nuove dimensioni per ogni ripetizione sono 1471×6.

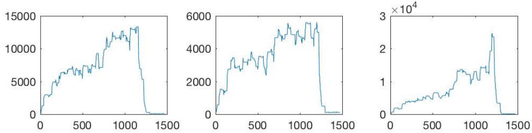


Figura: Posa 1, ripetizione 3, segnale preprocessato nei 6 sensori.

Riduzione della dimensionalità

Osservando gli autovalori della matrice di covarianza del dataset

$$\text{EIG} = [2.67 \cdot 10^{7}, 2.79 \cdot 10^{6}, 9.61 \cdot 10^{5}, 3.90 \cdot 10^{5}, 1.4 \cdot 10^{5}, 9.73 \cdot 10^{-24}]$$

ricaviamo l'informazione totale

$$\text{Tot} = \sum_{i=1}^{6} EIG(i).$$

L'informazione di ogni componente principale si ricava

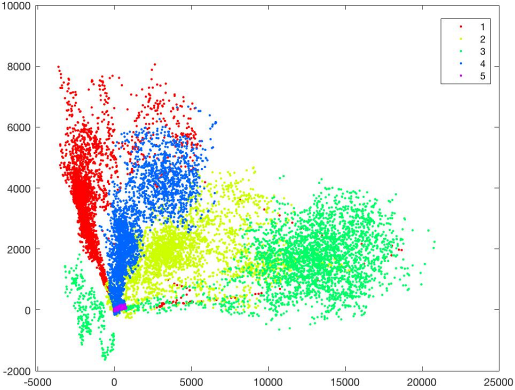
$$\text{Tot} : 100 = \text{EIG}(i) : \text{information}(i) \qquad \text{i=1}, \ldots, 6.$$

$$\Rightarrow \text{information} = [0.86, 0.09, 0.031, 0.012, 0.0045, 3.14 \cdot 10^{-31}].$$

Riduzione della dimensionalità

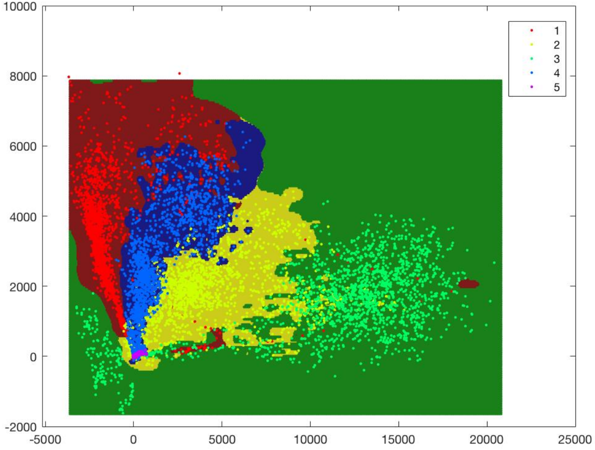
Siano y_1 e y_2 le prime due componenti principali

$$span\{y_1, y_2\} \quad \Rightarrow \quad 95\% \text{ di } \ \text{Tot} \ \text{(informazione totale).}$$



Classificazione: primo approccio

Algoritmo: SVM con kernel gaussiano
⇒ **Randomizzazione del dataset**



Classificazione: primo approccio

$$\begin{aligned} pos &= [4903, 4614, 5032, 4461, 4400], \\ falspos &= [72, 310, 251, 229, 801], \\ falsneg &= [217, 587, 140, 633, 23]. \end{aligned}$$

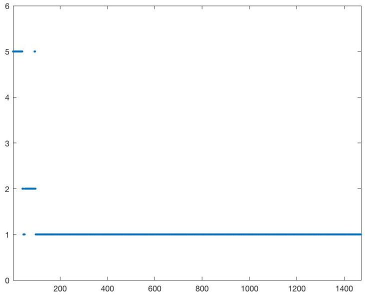
$$acc(i) = \frac{pos(i)}{pos(i) + falspos(i) + falsneg(i)} \qquad i = 1, \dots, 5.$$

class	accuracy
1	94%
2	84%
3	93%
4	84%
5	84%

Classificazione: secondo approccio

⇒ **Struttura matriciale delle ripetizioni viene mantenuta**

	1st interval	2nd interval	3rd interval	4th interval
gesture 1	6 (7%)	64 (80%)	80 (100%)	80(100%)
gesture 2	34 (43%)	13 (16%)	0	0
gesture 3	0	0	0	0
gesture 4	0	0	0	0
gesture 5	40 (50%)	3 (4%)	0	0
voting	abstension	1	1	1



Classificazione: secondo approccio

Fissata una posa, considero le $\binom{10}{6} = 210$ combinazioni di 10 elementi presi 6 alla volta (analoghe combinazioni per le altre pose).

- In ognuno di questi 210 casi seleziono le ripetizioni per il training, le complementari 4 sono usate per il test.
- Addestro il classificatore 210 volte e predico i risultati.
- Per ogni posa conto il numero di ripetizioni correttamente classificate e divido per il numero totale dei casi $210 \cdot 4$.

I risultati di tale statistica sono riportati nella tabella a sinistra. Nella tabella a destra invece sono riportati i risulati di corretta classificazione del transiente in uno dei primi due intervalli di 80 punti.

class	accuracy
1	90%
2	55%
3	90%
4	51%
5	99%

class	accuracy
1	91%
2	55%
3	88%
4	68%
5	100%