Support vector machines

Classificatori

- Classificatore mediante iperpiano separante
- Classificatore a margine massimo
- Classificatore a vettori di supporto
- Support Vector Machines

Iperpiano

Iperpiano: in uno spazio p-dimensionale l'iperpiano è un sottospazio planare di dimensione p-1. Per dimensioni maggiori di 3 è difficile visualizzare un iperpiano, ma fino a 3 dimensioni la sua definizione è immediata.

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0$$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p = 0$$

Quindi, se un punto $X = (X_1, \ldots, X_p)$ soddisfa le relazioni precedenti allora questo giace sull'iperpiano.

Iperpiano

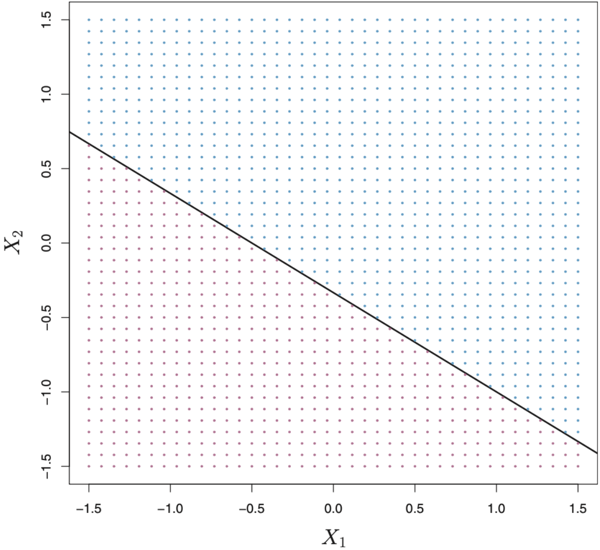
Se invece si ha che

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p > 0$$

o

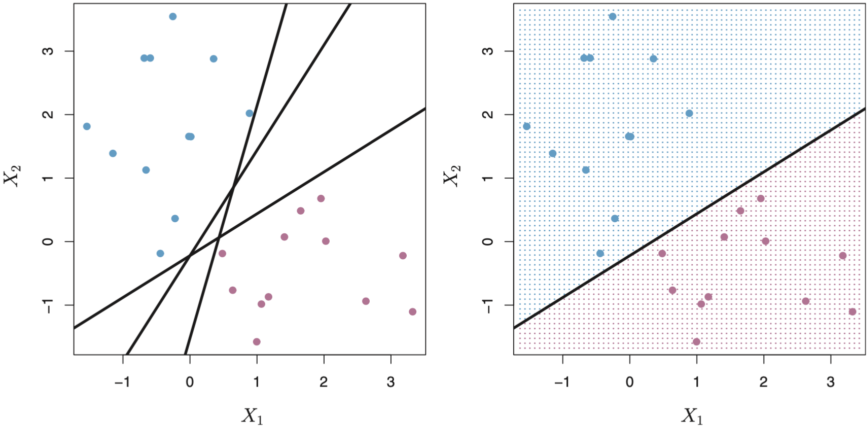
$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p < 0$$

allora il punto si troverà ad una parte o all'altra dell'iperpiano.



$$1 + 2X_1 + 3X_2 = 0$$

Classificatore mediante iperpiano separante



Classificatore mediante iperpiano separante

Indicando con $y_i = 1$ le osservazioni di classe blu e con $y_i = -1$ le osservazioni della classe viola si avrà:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p > 0 \quad se \quad y_i = 1$$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p < 0 \quad se \quad y_i = -1$$

o anche:

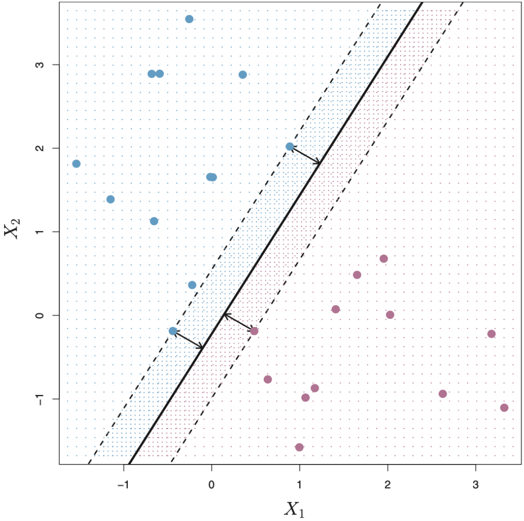
$$y_i(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p) > 0$$

Classificatore a margine massimo

Si sceglie, fra tutti i possibili, quell'iperpiano che ha la distanza minima più lontana dalle osservazioni di training.
La classificazione del test avviene in maniera semplice, considerando il segno di:

$$f(x^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \ldots + \beta_p x_p^*$$

Classificatore a margine massimo

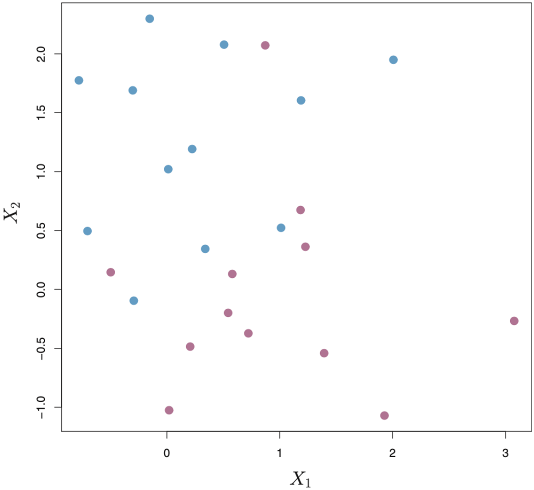


Classificatore a margine massimo

$$y_i(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p) > M$$

con vincolo

$$\sum_{j=1}^{p} \beta_j^2 = 1$$

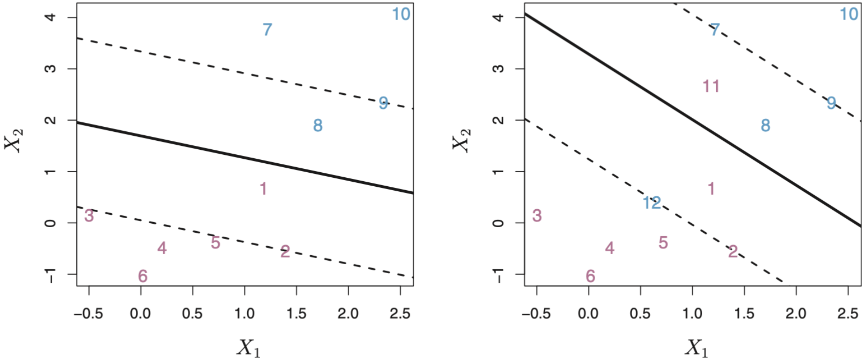

2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
X_2
0
1
2
3
X_1

Classificatore a margine morbido-Classificatore support vector

Classificatore che non separa perfettamente le due classi ma:

- Maggiore robustezza rispetto a singole osservazioni;
- No overfitting

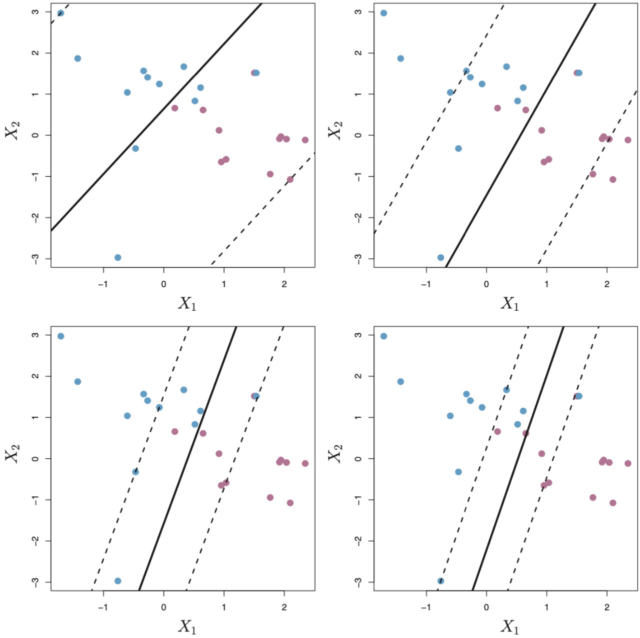
Si accetta che alcune osservazioni siano nel lato sbagliato del margine o anche nell'iperpiano sbagliato.

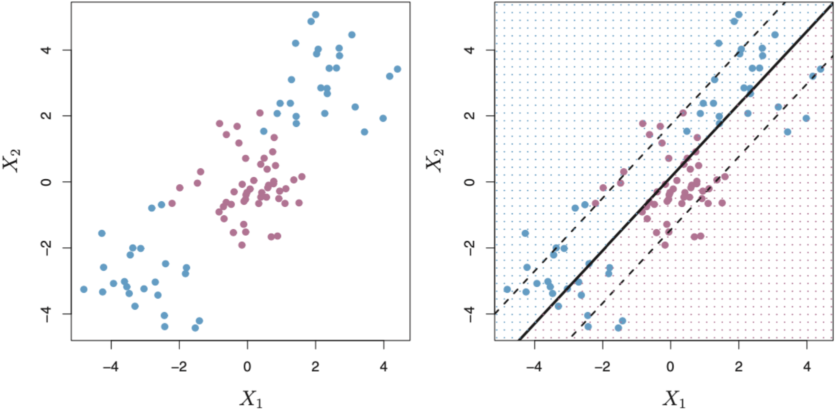


10
7
9
8
1
3
5
4
2
6
11
12
−1
0
1
2
3
4
X_2
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
X_1

$$y_i(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p) \geq M(1 - \varepsilon)$$

- $\varepsilon_i \geq 0$
- $\sum_{i=1}^{n} \varepsilon_i \leq C$
- se $\varepsilon_i = 0$: osservazione lato corretto iperpiano
- se $\varepsilon_i > 0$: osservazione dalla parte sbagliata del margine
- se $\varepsilon_i > 1$: osservazione dalla parte sbagliata dell'iperpiano
- C parametro di tuning che rappresenta quanto il margine può essere "violato"


X_1
X_2
−1
0
1
2
−3
−2
3



X_2
X_1
−4
−2
0
2
4

Classificazione con confini non lineari

Massimizzazione di M con:

$$y_i \left(\beta_0 + \sum_{j=1}^{p} \beta_{j1} x_{ij} + \sum_{j=1}^{p} \beta_{j2} x_{ij}^2 \right) \geq M(1 - \varepsilon)$$

- $\sum_{i=1}^{n} \varepsilon_i \leq C$
- $\varepsilon_i \geq 0$
- $\sum_{j=1}^{p} \sum_{k=1}^{2} \beta_{jk}^2 = 1$

Support vector machines

Ampliamento dello spazio delle variabili utilizzando i Kernel