Esercizi su SVM per classificazione binaria

Ricordiamo che, per un insieme linearmente separabile, la formulazione duale per una SVM binaria è la seguente:

$$\max_\alpha \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \vec{x}_i \cdot \vec{x}_j$$
$$\text{soggetto a: } \forall i \in \{1, \ldots, n\} : \alpha_i \geq 0 \text{ e } \sum_{i=1}^n y_i \alpha_i = 0.$$

I valori ottimi delle α_i^* determinano l'iperpiano ottimo secondo la formula:

$$\vec{w^*} = \sum_{i=1}^n y_i \alpha_i^* \vec{x}_i$$

a meno del valore di b che viene determinato ricordando che per un qualsiasi vettore di supporto $\vec{x}_s$ deve valere $\boxed{y_s(\vec{w}^* \cdot \vec{x}_s + b^*) = 1}$ e quindi considerando un esempio positivo $(y_s = 1)$

$$b^* = 1 - \vec{w}^* \cdot \vec{x}_s$$

Esercizio 1

Si consideri il seguente insieme di apprendimento con 3 esempi:

```
1: ([2,1], -1)
2: ([1,3], +1)
3: ([4,1], -1)
```

si calcoli il vettore $\vec{w^*}$ e la soglia b^* corrispondenti all'iperpiano ottimo separatore, cioè la soluzione restituita da una Support Vector Machine (senza utilizzo di kernel).

Soluzione:

Si noti che da una ispezione visuale degli esempi si riconosce subito che l'insieme di apprendimento è linearmente separabile. Pertanto, per trovare la soluzione usiamo la corrispondente formulazione duale per la SVM. In particolare, dobbiamo massimizzare, rispetto alle variabili duali, la funzione:

$$L(\vec{\alpha}) = \sum_{i=1}^{3} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{3} y_i y_j \alpha_i \alpha_j \vec{x}_i \cdot \vec{x}_j$$

dove $y_1 = y_3 = -1$, $y_2 = +1$, $\vec{x}_1 \equiv [2, 1]$, $\vec{x}_2 \equiv [1, 3]$, $\vec{x}_3 \equiv [4, 1]$, rispettando i vincoli

$$\alpha_1 \geq 0, \quad \alpha_2 \geq 0, \quad \alpha_3 \geq 0$$

e

$$\sum_{i=1}^{3} y_i \alpha_i = -\alpha_1 + \alpha_2 - \alpha_3 = 0.$$

Sapendo che,

$$\begin{aligned}
\vec{x}_1 \cdot \vec{x}_1 &= 5 \\
\vec{x}_2 \cdot \vec{x}_2 &= 10 \\
\vec{x}_3 \cdot \vec{x}_3 &= 17 \\
\vec{x}_1 \cdot \vec{x}_2 &= 5 \\
\vec{x}_1 \cdot \vec{x}_3 &= 9 \\
\vec{x}_2 \cdot \vec{x}_3 &= 7
\end{aligned}$$

la funzione lagrangiana può essere scritta come:

$$L(\vec{\alpha}) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2} \left(5\alpha_1^2 + 10\alpha_2^2 + 17\alpha_3^2 - 10\alpha_1\alpha_2 + 18\alpha_1\alpha_3 - 14\alpha_2\alpha_3\right)$$

Poiché deve vale il vincolo $\alpha_2 = \alpha_1 + \alpha_3$, possiamo eliminare da $L()$ la variabile α_2:

$$\begin{aligned}
L(\alpha_1, \alpha_3) &= 2\alpha_1 + 2\alpha_3 - \frac{1}{2} \left(5\alpha_1^2 + 10(\alpha_1 + \alpha_3)^2 + 17\alpha_3^2 + \right. \\
&\quad \left. -10\alpha_1(\alpha_1 + \alpha_3) + 18\alpha_1\alpha_3 - 14\alpha_3(\alpha_1 + \alpha_3)\right) \\
&= 2\alpha_1 + 2\alpha_3 - \frac{5}{2}\alpha_1^2 - \frac{13}{2}\alpha_3^2 - 7\alpha_1\alpha_3
\end{aligned}$$

Il massimo per la funzione $L(\alpha_1, \alpha_3)$ si ottiene ponendo le derivate prime rispetto a α_1 e α_3 uguali a 0:

$$\begin{aligned}
\frac{\partial L(\alpha_1, \alpha_3)}{\partial \alpha_1} &= 2 - 5\alpha_1 - 7\alpha_3 = 0 \\
\frac{\partial L(\alpha_1, \alpha_3)}{\partial \alpha_3} &= 2 - 13\alpha_3 - 7\alpha_1 = 0
\end{aligned}$$

Tale sistema lineare ha soluzione $\alpha_1 = \frac{3}{4} > 0$, $\alpha_3 = -\frac{1}{4} < 0$ (!). Poiché $\alpha_3 < 0$ il vincolo di non-negatività è violato. Si noti che il valore di $\alpha_3 \geq 0$ per cui si ottiene il valore massimo di $L(\alpha_1, \alpha_3)$ è 0, perché $L(\alpha_1, \alpha_3)$ è convessa con massimo a sinistra del valore 0. Quindi, fissato $\alpha_3^* = 0$, il valore ottimo per α_1 si ottiene massimizzando la funzione $L(\alpha_1, 0)$, cioè risolvendo l'equazione:

$$\frac{\partial L(\alpha_1, 0)}{\partial \alpha_1} = 2 - 5\alpha_1 = 0$$

che ha soluzione ottima $\alpha_1^* = \frac{2}{5} > 0$, e quindi, poiché $\alpha_2 = \alpha_1 + \alpha_3$, $\alpha_2^* = \frac{2}{5} > 0$. Si noti che, in questo caso, solo gli esempi 1 e 2 sono di supporto. Il vettore dei pesi ottimo è quindi

$$\vec{w}^* = \sum_{i=1}^{2} y_i \alpha_i^* \vec{x}_i = [-\frac{2}{5}, \frac{4}{5}]$$

Infine, calcoliamo il valore ottimo di b utilizzando il vettore di supporto 2:

$$b^* = 1 - \vec{w}^* \cdot \vec{x}_2 = -1$$

Esercizio 2

Si consideri il seguente insieme di apprendimento con 4 esempi:

```
1: ([1,1], +1)
2: ([2,2], +1)
3: ([4,2], -1)
4: ([2,3], -1)
```

si calcoli il vettore $\vec{w}^*$ e la soglia b^* corrispondenti all'iperpiano ottimo separatore, cioè la soluzione restituita da una Support Vector Machine (senza utilizzo di kernel). Si dica anche a quanto ammonta il margine separatore. (Aiuto: usate la rappresentazione grafica dell'insieme di apprendimento per riconoscere immediatamente il/i vettore/i che non è/sono di supporto).

Soluzione:

Si noti che da una ispezione visuale degli esempi si riconosce subito che l'insieme di apprendimento è linearmente separabile. Inoltre risulta subito chiaro che l'esempio 1 non è di supporto perché "coperto" dall'esempio 2, che risulterà sicuramente essere strettamente più vicino all'iperpiano ottimo di quanto lo sarà l'esempio 1. Quindi possiamo già imporre $\alpha_1^* = 0$.

Per trovare la soluzione dobbiamo massimizzare, rispetto alle variabili duali, la funzione:

$$L(\vec{\alpha}) = \sum_{i=1}^{4} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{4} y_i y_j \alpha_i \alpha_j \vec{x}_i \cdot \vec{x}_j$$

dove $y_1 = +1$, $y_2 = +1$, $y_3 = -1$, $y_4 = -1$, $\vec{x}_1 \equiv [1,1]$, $\vec{x}_2 \equiv [2,2]$, $\vec{x}_3 \equiv [4,2]$, $\vec{x}_4 \equiv [2,3]$, rispettando i vincoli

$$\alpha_1 \geq 0,\ \ \alpha_2 \geq 0,\ \ \ \alpha_3 \geq 0,\ \ \ \alpha_4 \geq 0$$

e

$$\sum_{i=1}^{4} y_i \alpha_i = \alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0.$$

Sapendo che $\alpha_1^* = 0$, le informazioni di cui abbiamo bisogno per procedere sono

$$\begin{aligned}
\vec{x}_2 \cdot \vec{x}_2 &= 8 \\
\vec{x}_3 \cdot \vec{x}_3 &= 20 \\
\vec{x}_4 \cdot \vec{x}_4 &= 13 \\
\vec{x}_2 \cdot \vec{x}_3 &= 12 \\
\vec{x}_2 \cdot \vec{x}_4 &= 10 \\
\vec{x}_3 \cdot \vec{x}_4 &= 14
\end{aligned}$$

la funzione lagrangiana, (sapendo che $\alpha_1^* = 0$) può essere scritta come:

$$L(\vec{\alpha}) = \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}\left(8\alpha_2^2 + 20\alpha_3^2 + 13\alpha_4^2 - 24\alpha_2\alpha_3 - 20\alpha_2\alpha_4 + 28\alpha_3\alpha_4\right)$$

Poiché deve vale il vincolo $\alpha_2 = \alpha_3+\alpha_4$, possiamo eliminare da $L()$ la variabile α_2:

$$
\begin{aligned}
L(\alpha_3, \alpha_4) &= 2\alpha_3 + 2\alpha_4 - \frac{1}{2}\left(8(\alpha_3+\alpha_4)^2 + 20\alpha_3^2 + 13\alpha_4^2 - 24(\alpha_3+\alpha_4)\alpha_3\right. \\
&\quad \left. -20(\alpha_3+\alpha_4)\alpha_4 + 28\alpha_3\alpha_4\right) \\
&= 2\alpha_3 + 2\alpha_4 - 2\alpha_3^2 - \frac{1}{2}\alpha_4^2
\end{aligned}
$$

Il massimo per la funzione $L(\alpha_3, \alpha_4)$ si ottiene ponendo le derivate prime rispetto a α_3 e α_4 uguali a 0:

$$
\begin{aligned}
\frac{\partial L(\alpha_3, \alpha_4)}{\partial \alpha_3} &= 2 - 4\alpha_3 = 0 \\
\frac{\partial L(\alpha_3, \alpha_4)}{\partial \alpha_4} &= 2 - \alpha_4 = 0
\end{aligned}
$$

Tale sistema lineare ha soluzione $\alpha_3^* = \frac{1}{2} > 0$, $\alpha_4^* = 2 > 0$, e usando l'equazione $\alpha_2 = \alpha_3 + \alpha_4$ otteniamo $\alpha_2^* = \frac{5}{2} > 0$.

Il vettore dei pesi ottimo è quindi

$$
\vec{w}^* = \sum_{i=1}^{4} y_i \alpha_i^* \vec{x}_i = [-1, -2]
$$

Infine, calcoliamo il valore ottimo di b utilizzando il vettore di supporto 2:

$$
b^* = 1 - \vec{w}^* \cdot \vec{x}_1 = 7
$$

Infine, il margine di separazione sarà uguale a $\frac{2}{||\vec{w}^*||} = \frac{2}{\sqrt{5}}$