

Apprendimento Automatico

Fabio Aiolli
www.math.unipd.it/~aiolli

Sito web del corso
www.math.unipd.it/~aiolli/corsi/1516/aa/aa.html

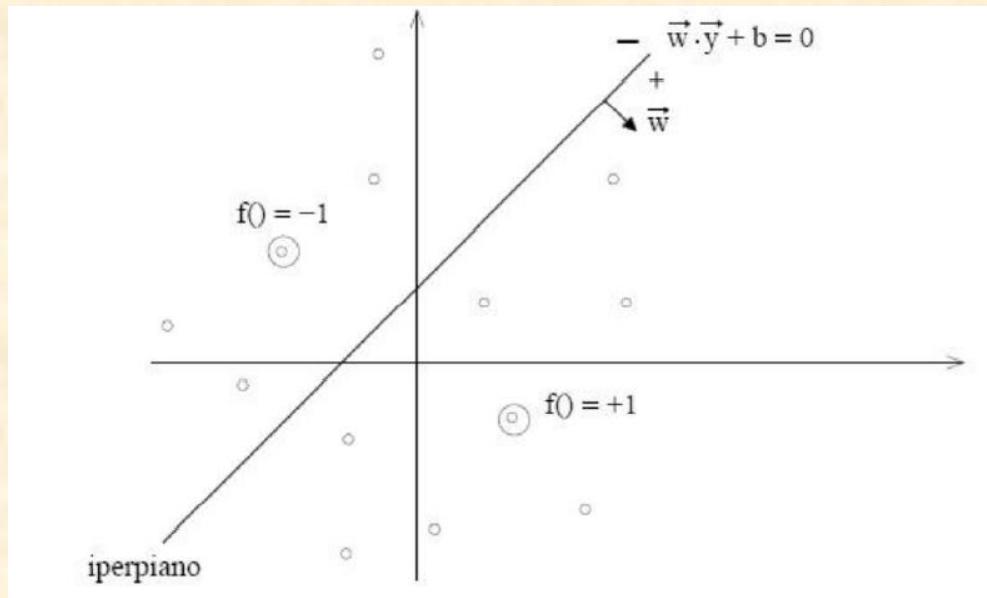
Spazio delle ipotesi

Esempio 1

Iperpiani in R^2

- Spazio delle istanze: punti nel piano $X = \{y \mid y \in R^2\}$
- Spazio delle ipotesi: dicotomie indotte da iperpiani in R^2

$$H = \{f_{(w,b)}(y) \mid f_{(w,b)}(y) = \text{sign}(w \cdot y - b), w \in R^2, b \in R\}$$



cosa cambia per R^n ?

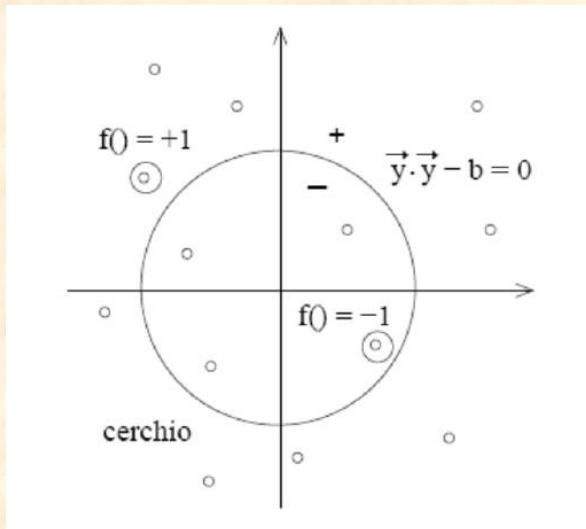
Spazio delle ipotesi

Esempio 2

Dischi in R^2

- Spazio delle istanze: punti nel piano $X = \{y \mid y \in R^2\}$
- Spazio delle ipotesi: dicotomie indotte da dischi in R^2 centrati nell'origine

$$H = \{f_b(y) \mid f_b(y) = \text{sign}(\|y\|^2 - b), b \in R\}$$



cosa cambia per R^n ?

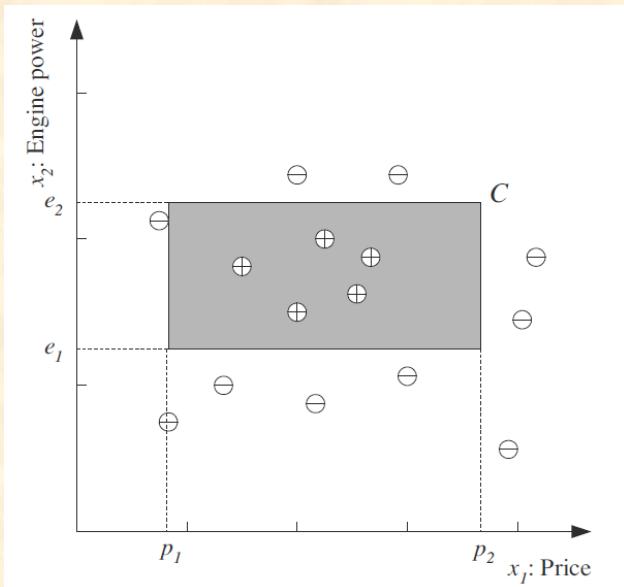
Spazio delle ipotesi

Esempio 3

Rettangoli in \mathbb{R}^2

- Spazio delle istanze: punti nel piano $X = \{x | x \in \mathbb{R}^2\}$
- Spazio delle ipotesi: dicotomie indotte da rettangoli in \mathbb{R}^2

$$H = \{f_\theta(y) | f_\theta(y) = p_1 \leq x_1 \leq p_2 \cap e_1 \leq x_2 \leq e_2, \theta = \{p_1, p_2, e_1, e_2\}\}$$



cosa cambia per \mathbb{R}^n ?

Spazio delle ipotesi

Esempio 4

Congiunzione di m letterali positivi

- Spazio delle istanze: stringhe di m bits: $X = \{s \mid s \in \{0,1\}^m\}$
- Spazio delle ipotesi: tutte le sentenze logiche che riguardano letterali positivi l_1, \dots, l_m (l_1 vero se il primo bit vale 1, l_2 vero se il secondo bit vale 1, ecc.) e che contengono solo operatore and.

$$H = \left\{ f_{\{i_1, \dots, i_j\}}(s) \mid f_{\{i_1, \dots, i_j\}}(s) \equiv l_{i_1} \cap l_{i_2} \cap \dots \cap l_{i_j}, \{i_1, \dots, i_j\} \subseteq \{1, \dots, m\} \right\}$$

Es. $m=3$, $X=\{0,1\}^3$

Alcune istanze: $s_1=101$, $s_2=001$, $s_3=100$, $s_4=111$

Alcune ipotesi: $h_1=l_2$, $h_2=l_1 \cap l_2$, $h_3=\text{true}$, $h_4=l_1 \cap l_3$, $h_5=l_1 \cap l_2 \cap l_3$

Notare che h_1, h_2, h_5 sono false per s_1, s_2, s_3 e vere per s_4 ; h_3 è vera per ogni istanza; h_4 è vera per s_1, s_4 ma falsa per s_2 e s_3 ;

Misurare la complessità dello spazio delle ipotesi: VC dimension

Frammentazione (Shattering)

Dato $S \subseteq X$, S è frammentato (shattered) dallo spazio delle ipotesi H se:

$\forall S' \subseteq S, \exists h \in H$, tale che $\forall x \in S, h(x) = 1 \Leftrightarrow x \in S'$

(ovvero H realizza tutte le possibili dicotomie di S)

VC (Vapnik-Chervonenkis) Dimension

La VC dimension di uno spazio delle ipotesi H definito su uno spazio delle istanze X è data dalla cardinalità del sottoinsieme più grande frammentato da H .

$$VC(H) = \max_{S \subseteq X} |S|: H \text{ frammenta } S$$

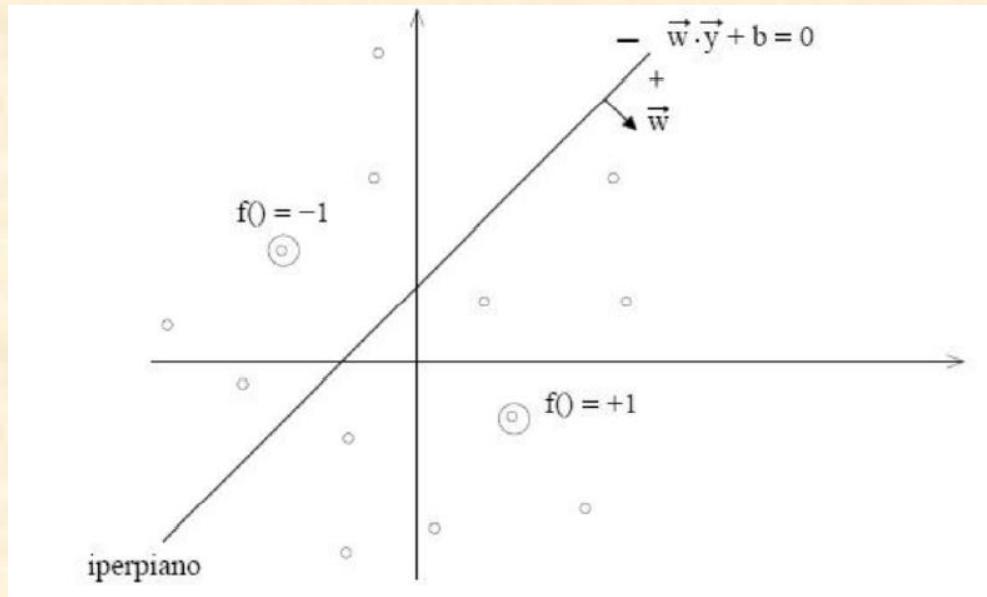
$$VC(H) = \infty \text{ se } S \text{ non è limitato}$$

VC dimension Esempio 1

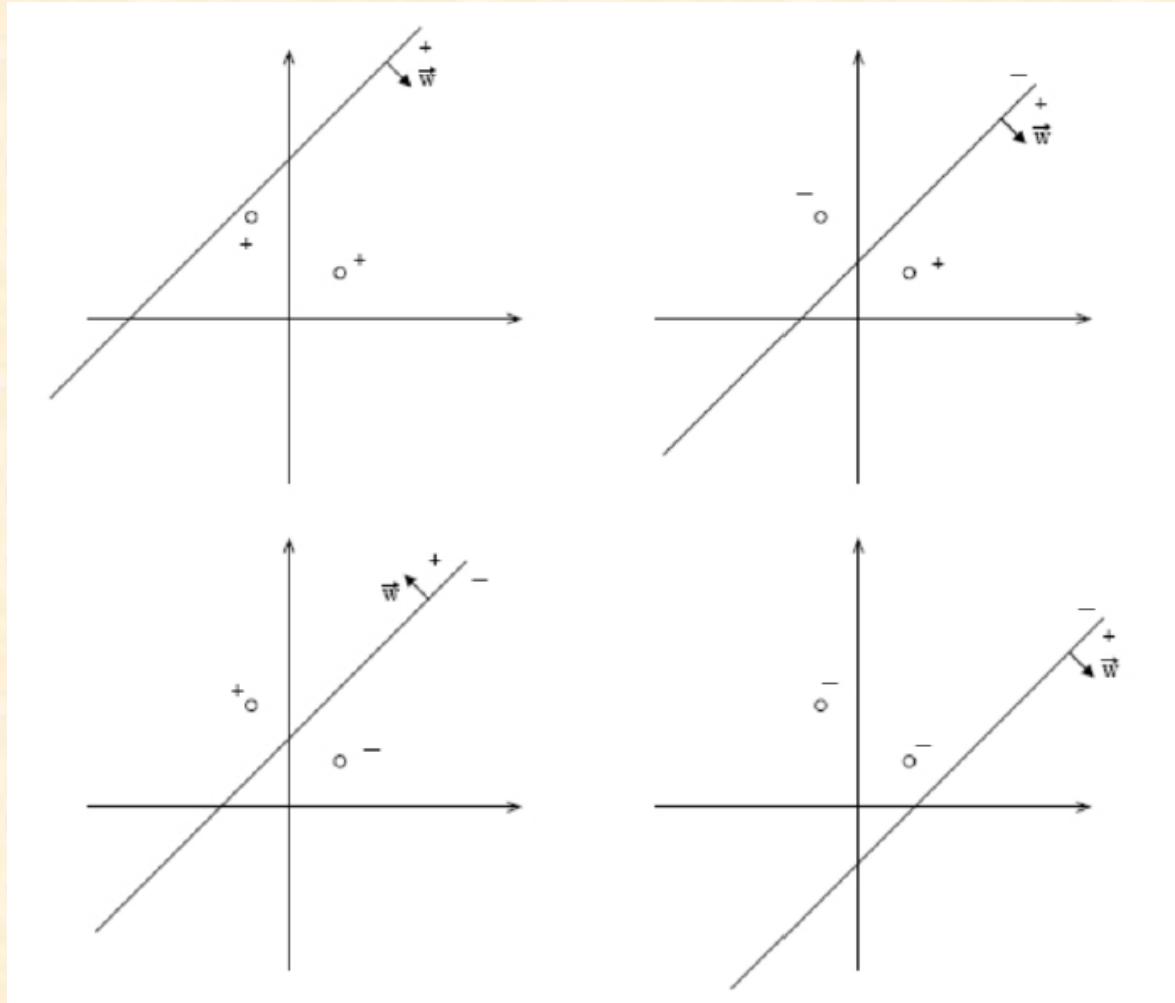
Iperpiani in R^2

- Spazio delle istanze: punti nel piano $X = \{y \mid y \in R^2\}$
- Spazio delle ipotesi: dicotomie indotte da iperpiani in R^2

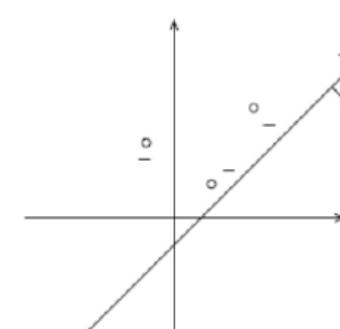
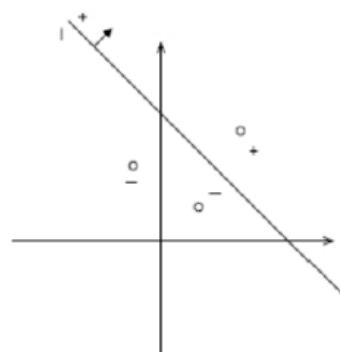
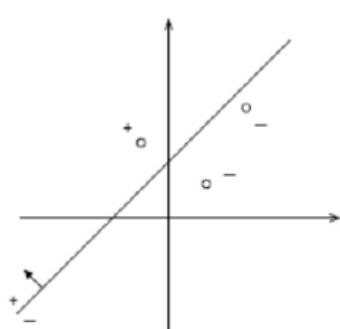
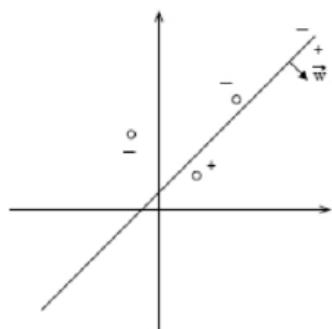
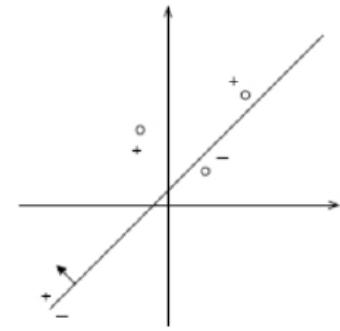
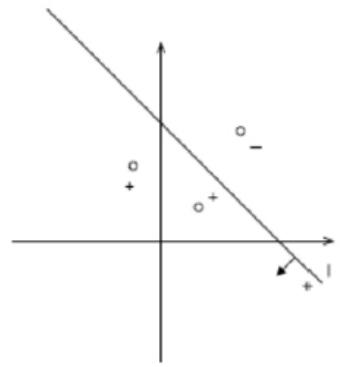
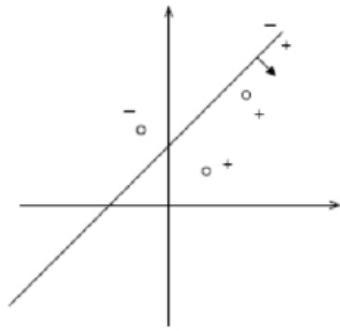
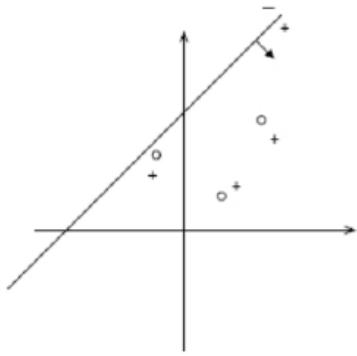
$$H = \{f_{(w,b)}(y) \mid f_{(w,b)}(y) = \text{sign}(w \cdot y + b), w \in R^2, b \in R\}$$



VC dimension Esempio 1 (2 punti)



VC dimension Esempio 1 (3 punti)

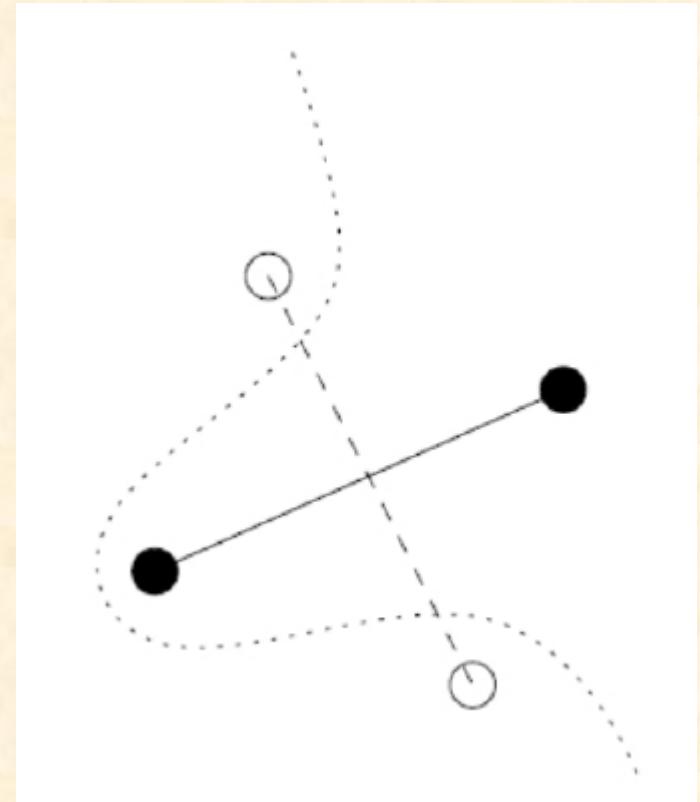


VC dimension Esempio 1 (4 punti)

Per 4 punti?

Esisteranno sempre due coppie di punti tali che, se unite con un segmento, provocano una intersezione tra i due segmenti e quindi, ponendo ogni coppia di punti in classi diverse, per separarli non basta una retta ma occorre una curva.

Quindi: $VC(H)=3$



Bound sull'errore di generalizzazione (classificazione binaria)

Considerando un problema di apprendimento binario:

Training set $S = \{(x_i, y_i)\}_{i=1, \dots, N}$

Spazio delle ipotesi $H = \{h_\theta(x)\}$

Algoritmo di apprendimento L che restituisce l'ipotesi $h_{\theta^*}(x)$ che minimizza l'errore empirico su S , $\text{errore}_S(h_\theta(x))$.

È possibile derivare un bound (limite superiore) all'errore ideale o errore di generalizzazione, valido con probabilità $(1-\delta)$ con δ piccolo a piacere, della forma

$$\text{errore}_D(h_\theta(x)) = \text{errore}_S(h_\theta(x)) + g(N, \text{VC}(H), \delta)$$

Bound sull'errore di generalizzazione (classificazione binaria)

Prendiamo i due termini del bound:

A = errore_S($h_{\theta}(x)$)

B = g(N, VC(H), δ)

Il **termine A** dipende dall'ipotesi restituita dall'algoritmo di apprendimento L.

Il **termine B** non dipende da L. Dipende dal numero di esempi di training (inversamente), dalla VC dimension dello spazio delle ipotesi e dalla confidenza.

Il termine B è spesso chiamato **VC-confidence** e risulta essere monotono rispetto al rapporto VC(H)/N. Fissato N, aumenta all'aumentare di VC(H).

Structural Risk Minimization (SRM)

L'approccio **Structural Risk Minimization (SRM)** proposto da Vapnik tenta di trovare un compromesso tra i due termini A e B visti sopra

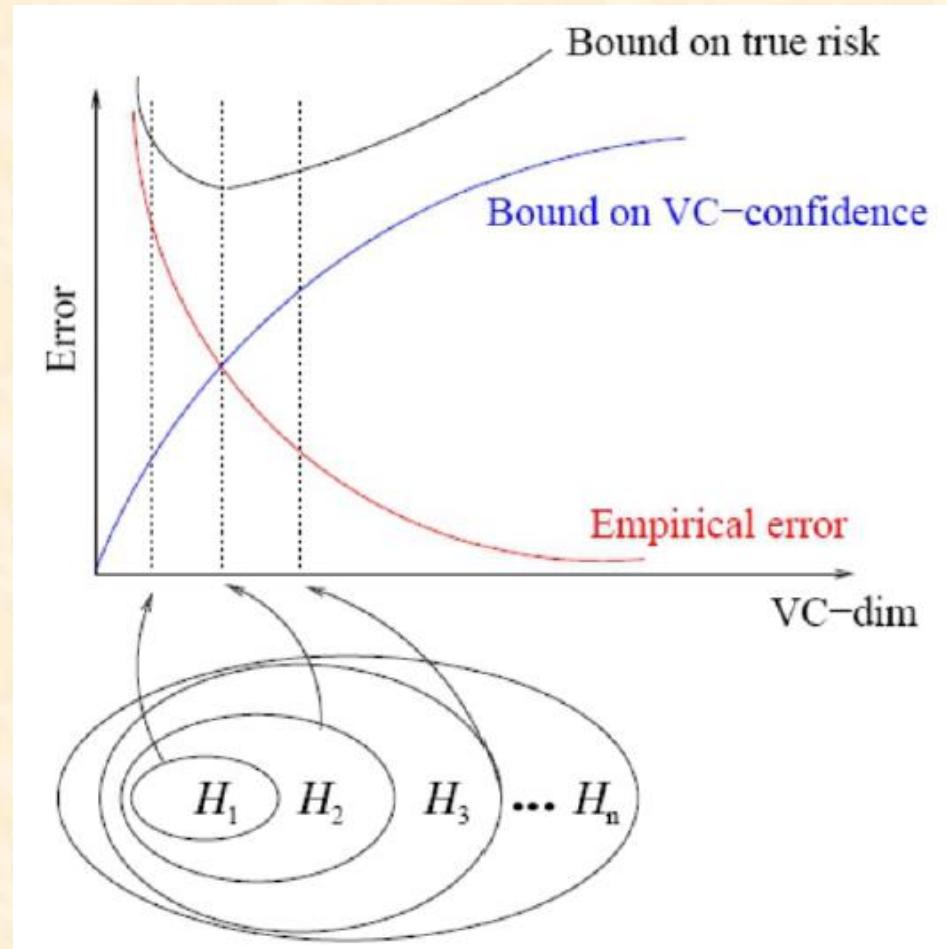
Si considerano:

$$H_1 \subseteq H_2 \subseteq \dots \subseteq H_n$$

tali che

$$VC(H_1) \leq VC(H_2) \leq \dots \leq VC(H_n)$$

Si seleziona l'ipotesi H_i che ha il valore del bound sull'errore di generalizzazione più piccolo.



Spazio delle ipotesi finito

Se lo spazio delle ipotesi è di cardinalità finita, allora:

$$VC(H) \leq \log_2(|H|)$$

Dimostriamolo:

Per ogni S frammentata da H , abbiamo $|H| \geq 2^{|S|}$, poiché H può realizzare tutte le possibili dicotomie di S , che sono esattamente $2^{|S|}$.

Scegliendo un S tale che $|S|=VC(H)$, si ottiene $|H| \geq 2^{|VC(H)|}$.

Applicando poi il logaritmo in base 2 su ambo i lati della disequazione ottenuta nel passo precedente, si ottiene il claim.