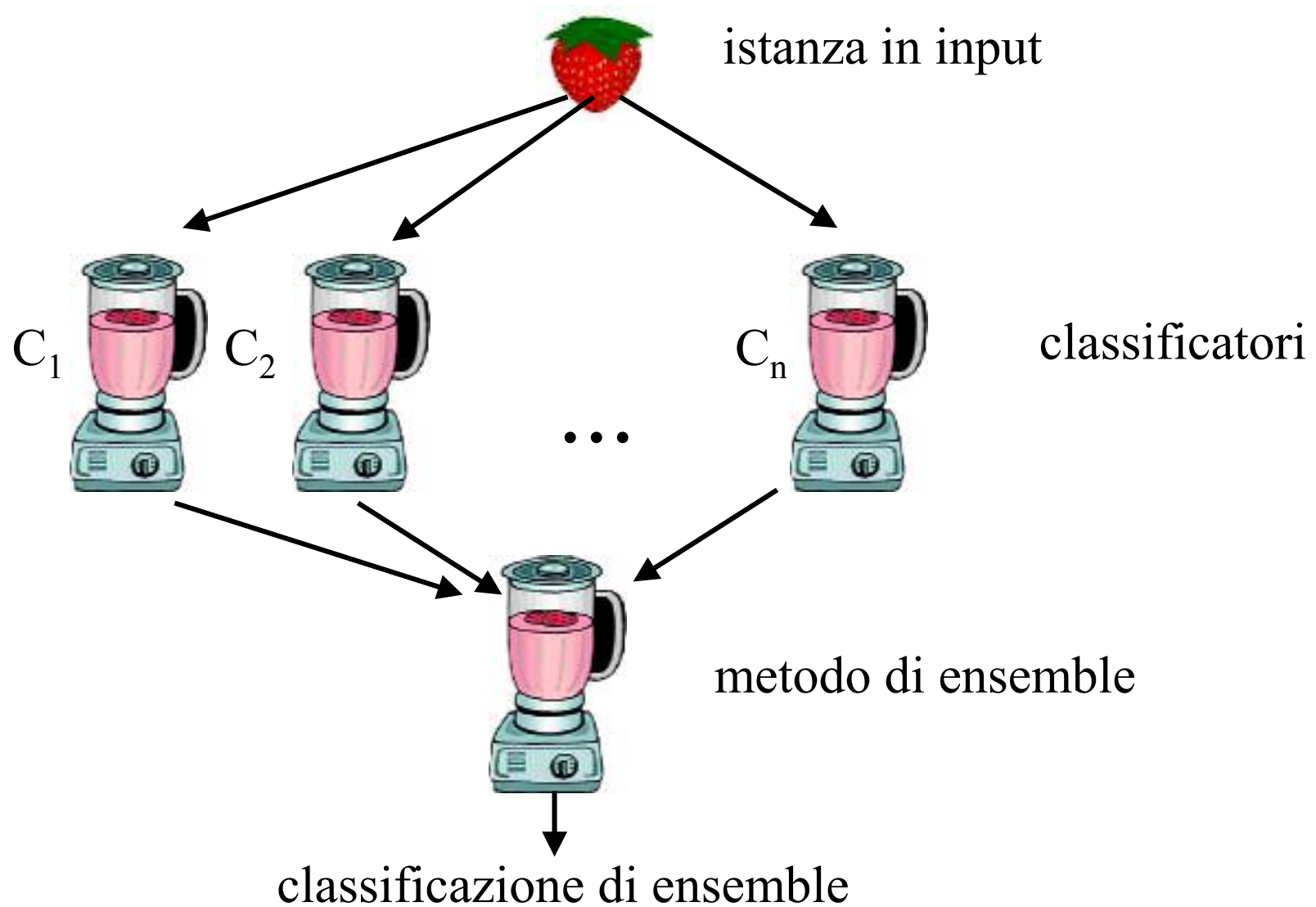


Metodi di Ensemble

Metaclassificatori

Metodi di Ensemble



Come combinare i classificatori?

- Normalmente mediante una media pesata (supponiamo che le classi siano $\{-1, +1\}$):

$$ensemble(x) = \text{sgn}\left(\sum_{i=1}^n w_i C_i(x)\right)$$

- w_i è il peso del classificatore C_i
- Si può estendere a un numero arbitrario classi
- **Voting semplice** ($w_k = w_j$ per ogni k, j)
- **Voting pesato** (i pesi influenzano la scelta finale)

Metodi di Ensemble

- Mediano i risultati di modelli differenti (o stesso modello parametrizzato in modo differente)
- Perché?
 - Normalmente hanno prestazioni migliori rispetto ai singoli classificatori
 - Più resistenti al rumore
- Perché no?
 - Richiedono più tempo
 - Overfitting

Bagging

- Dato un insieme di addestramento D , generiamo n insiemi di addestramento D_i ($|D_i| \leq |D|$)
- Ciascun insieme di addestramento D_i viene generato campionando esempi da D in modo uniforme (estratti con rimpiazzo)
 - *Bootstrapping*
- Addestra n classificatori sugli n insiemi di addestramento ottenuti
- Media i risultati sui vari classificatori (se l'output è reale) oppure applica una tecnica di voting (se l'output è basato su classi)

Bagging: Esempio (Opitz, 1999)

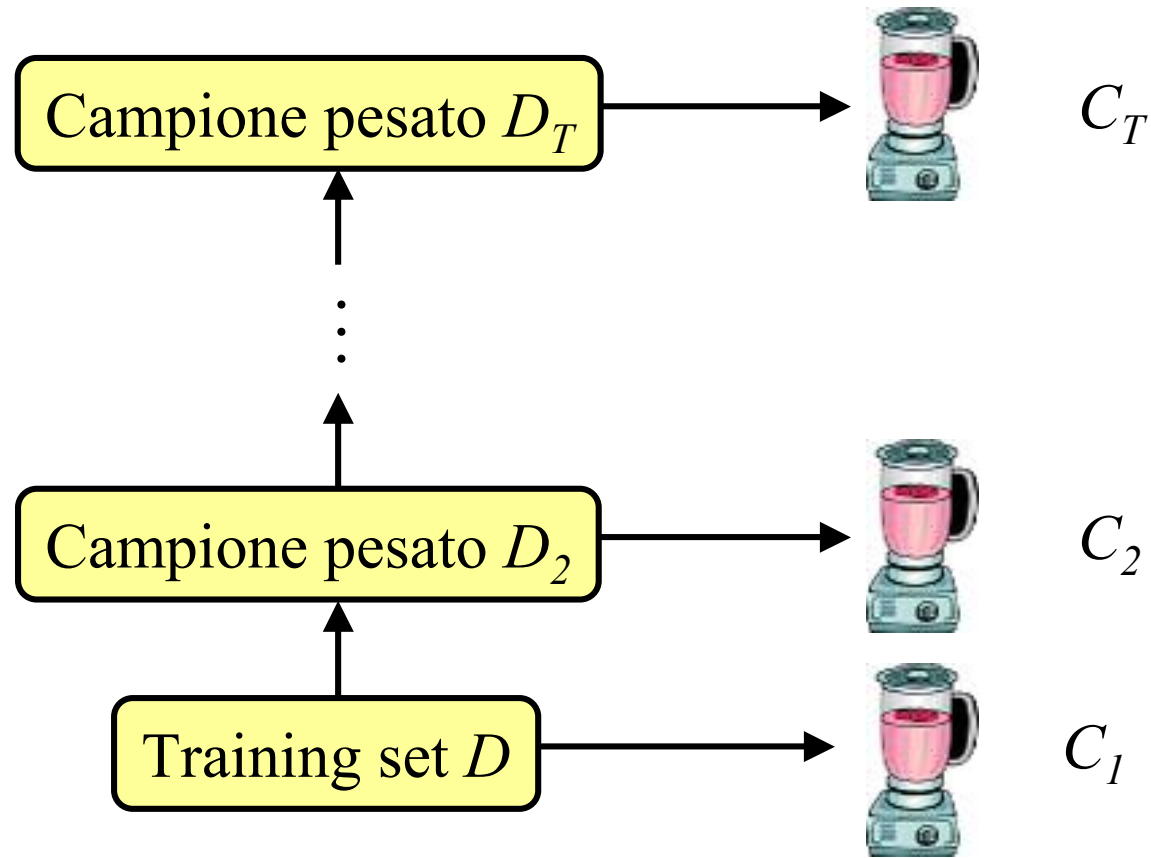
Insieme D	1	2	3	4	5	6	7	8
D_1	2	7	8	3	7	6	3	1
D_2	7	8	5	6	4	2	7	1
D_3	3	6	2	7	5	6	2	2
D_4	4	5	1	4	6	4	3	8

Boosting

- Può un insieme di classificatori “**deboli**” formare un classificatore “**forte**”?
- Il Boosting fornisce una risposta
- I classificatori vengono prodotti in sequenza
- Ciascun classificatore dipende dal precedente e tenta di migliorarne gli errori
- Gli esempi classificati in modo erraneo dai classificatori precedenti sono scelti più spesso o pesati maggiormente

Boosting

$$H(x) = \sum_{k=1}^T \alpha_k C_k(x)$$



Adaptive Boosting (AdaBoost)

- Si sceglie una distribuzione iniziale di selezione degli esempi (x_i, y_i) , $D_1(i) = 1/m$ dato $|D|=m$ e per $i = 1, \dots, m$

- **For** $k = 1, \dots, T$

- Definisce ε_k come la somma delle probabilità per le istanze misclassificate dai classificatori precedenti

$$\varepsilon_k = \sum_{i=1}^m D_k(i)[y_i \neq C_k(x_i)]$$

- Apprendi un classificatore C_k che minimizzi l'errore ε_k (esci se $\varepsilon_k \geq 0.5$)

- Calcola un peso α_k del classificatore C_k :

$$\alpha_k = \frac{1}{2} \log \frac{1 - \varepsilon_k}{\varepsilon_k}$$

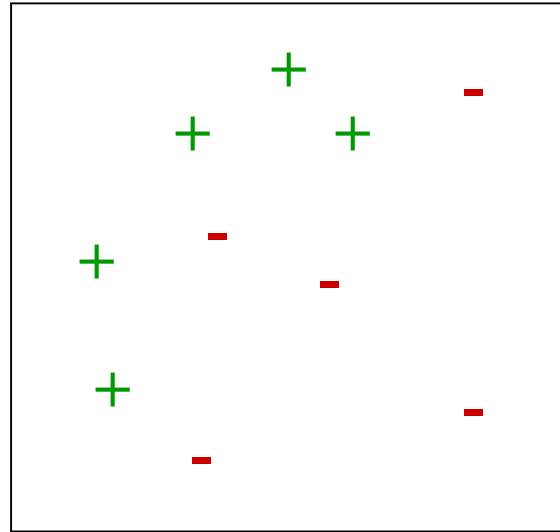
- Aggiorna la distribuzione:

- $D_{k+1}(i) = D_k(i)e^{\alpha_k}$ se x_i è classificato in modo **errato**,
- $D_{k+1}(i) = D_k(i)e^{-\alpha_k}$ se x_i è classificato in modo **corretto**

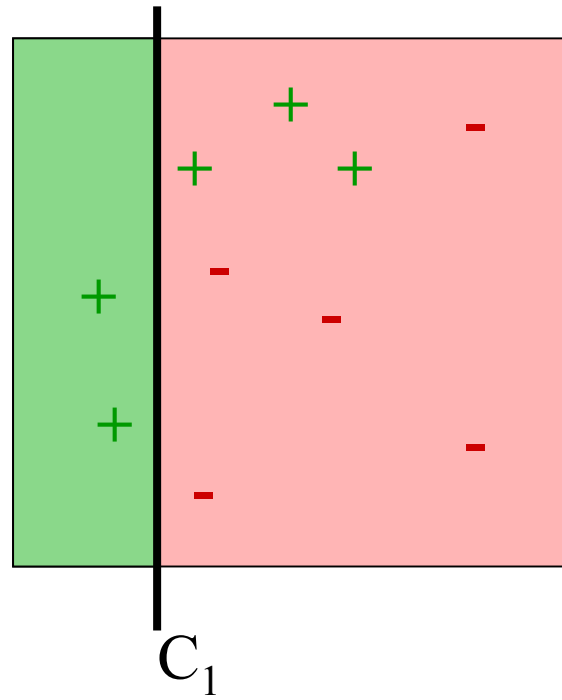
- “Rinormalizza” le probabilità (affinché sommino a 1)

- Combina i classificatori C_1, \dots, C_k usando il voting pesato

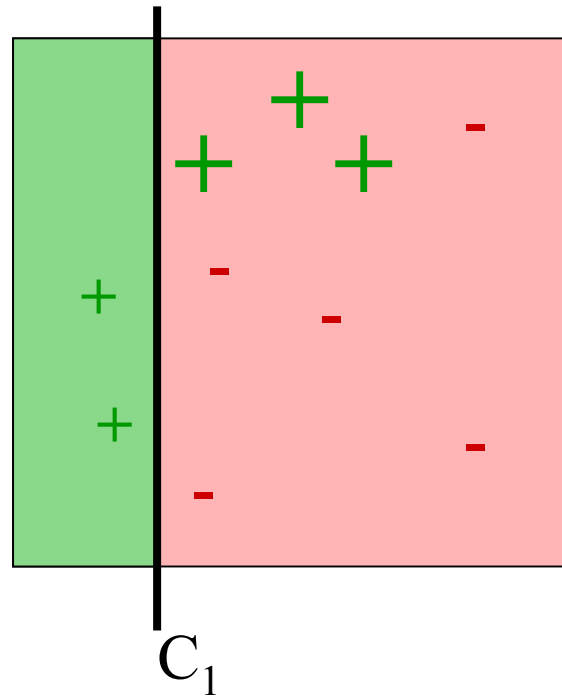
AdaBoost: esempio



AdaBoost: esempio (k=1)



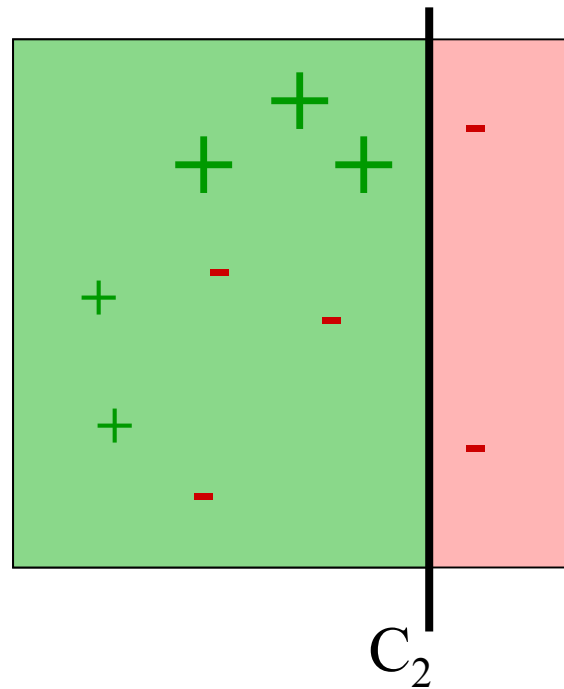
AdaBoost: esempio (k=1)



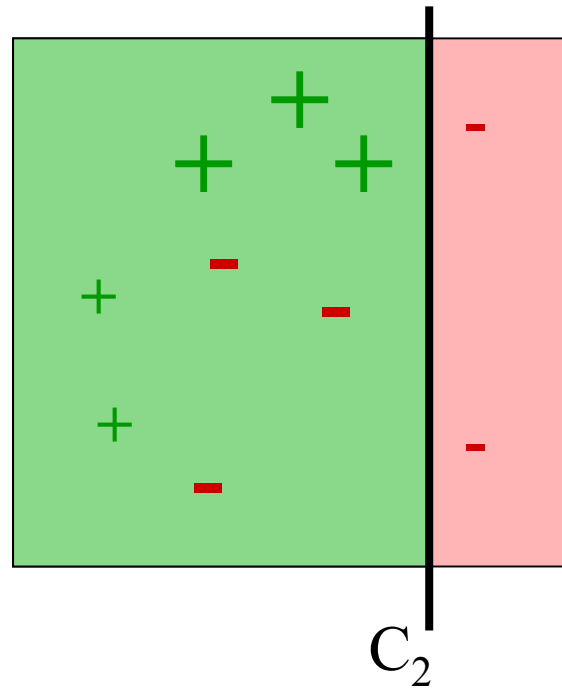
$$\varepsilon_1 = 0.30$$

$$\alpha_1 = 0.42$$

AdaBoost: esempio (k=2)



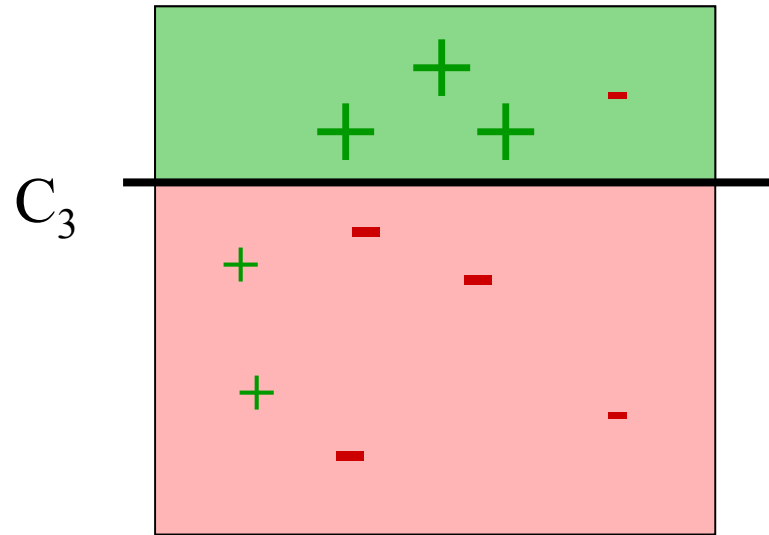
AdaBoost: esempio (k=2)



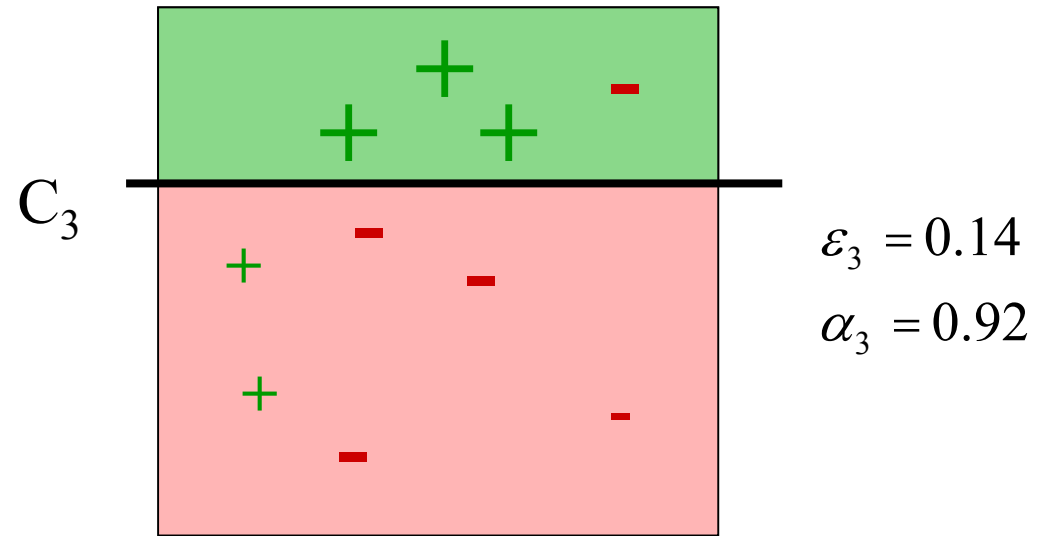
$$\varepsilon_2 = 0.21$$

$$\alpha_2 = 0.65$$

AdaBoost: esempio (k=3)



AdaBoost: esempio (k=3)



AdaBoost: H finale (T=3)

$$H(x) = \text{sgn}(0.42 \cdot \text{[diagram 1]} + 0.65 \cdot \text{[diagram 2]} + 0.92 \cdot \text{[diagram 3]})$$

The diagram illustrates the final weak classifier $H(x)$ as a weighted sum of three weak classifiers. Each weak classifier is represented by a square with a vertical line and a dot in the center. The first weak classifier has a vertical line on the left, with the area to the left of the line shaded green and the area to the right shaded red. The second weak classifier has a vertical line on the right, with the area to the left of the line shaded green and the area to the right shaded red. The third weak classifier has a horizontal line at the top, with the area below the line shaded red and the area above shaded green.

AdaBoost

- Il classificatore finale è:

$$H(x) = \sum_{k=1}^T \alpha_k C_k(x)$$

- Vantaggi:
 - Solo un parametro da apprendere: T
 - Può essere combinato con qualsiasi classificatore “debole”
 - Garanzie teoriche data una quantità sufficiente di dati e un buon classificatore “debole”
- Svantaggi:
 - Suscettibile al rumore
 - Dipende dal classificatore “debole” e dai dati
- Provate anche questa applet:
 - <http://www.cse.ucsd.edu/~yfreund/adaboost/index.html>