

Approccio statistico alla classificazione

Approccio parametrico e non
parametrico

Finestra di Parzen

Classificatori K-NN 1-NN



Limitazioni dell'approccio bayesiano

- Con l'approccio bayesiano, sarebbe possibile costruire un classificatore ottimo se si conoscessero:
 - le probabilità a priori $P(\omega_i)$
 - le densità condizionate alla classe $P(x | \omega_i)$
- Informazioni che raramente sono disponibili
- Alternativa: costruire un classificatore da un insieme di esempi (training set)
 - Pro: stima delle $P(\omega_i)$ semplicemente realizzabile
 - Contro: training set troppo limitato per una stima affidabile delle distribuzioni condizionate





Approccio parametrico

- In questo tipo di approccio si assume nota la forma delle densità condizionali; tipicamente si assume una gaussiana $P(x | \omega_i) \sim N(\mu_i, \Sigma_i)$.
- Ci sono quindi due parametri da stimare per ogni classe.
- Le tecniche più usate per la stima sono:
 - Maximum-Likelihood (ML)
 - Stima Bayesiana
- Sebbene differenti nella logica, le due tecniche portano a risultati quasi identici.



Approccio non parametrico

- Nell'approccio parametrico tutte le densità erano unimodali (hanno un singolo massimo locale), mentre in molti problemi pratici le densità sono multimodali.
- Con l'approccio non parametrico si rimuove l'assunzione della conoscenza delle densità per cui si può lavorare con distribuzioni di forma arbitraria.
- Due tipologie di metodi non parametrici:
 - Stimare $p(x | \omega_j)$
 - Stimare direttamente le probabilità a posteriori $P(\omega_j|x)$



Stima della densità

- Consideriamo la probabilità che un vettore x , la cui densità è $p(x)$, cada in una regione R :

$$P = \int_R p(\xi) d\xi$$

- Consideriamo n campioni i.i.d. di x x_1, \dots, x_n . La probabilità che k di questi cadranno in R sarà data da:

$$P_k = \binom{n}{k} P^k (1-P)^{n-k}$$

mentre il valore atteso per k è $E[k]=nP$.



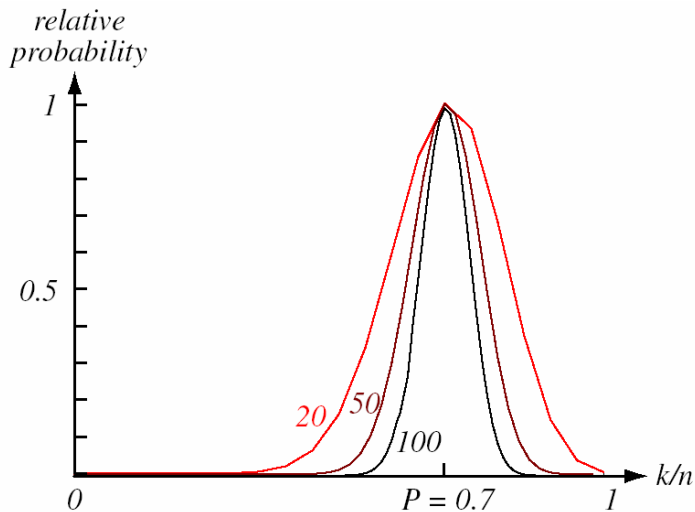
Stima della densità

- Possiamo quindi assumere $P \cong k/n$, dove la stima diventa più accurata al crescere di n .
- D'altra parte, se supponiamo $p(x)$ continua ed R sufficientemente piccola, possiamo porre:

$$P = \int_R p(\xi) d\xi \cong p(x) \int_R d\xi = p(x)V$$

- In questo modo, otteniamo una stima di $p(x)$:

$$p(x) \cong \frac{k/n}{V}$$



Stima della densità



Alcuni problemi:

- se fissassimo il volume V e facessimo crescere n , otterremmo una media di $p(x)$:

$$\frac{P}{V} = \frac{\int_R p(\xi) d\xi}{\int_R d\xi}$$

per cui dovremmo considerare un volume che tende a zero.

- Tuttavia, per n fissato, R diventerebbe talmente piccola che $k=0$ (e quindi $p(x) \cong 0$); altrimenti se $k>0$, la stima divergerebbe.



Stima della densità

- Mettiamoci nell'ipotesi di avere un numero illimitato di campioni.
- Per valutare $p(x)$ consideriamo una sequenza di regioni R_1, R_2, \dots, R_n contenenti x : la regione R_s si impiega nel caso $n=s$.
- Se V_n è il volume di R_n , k_n il numero di campioni che cadono in R_n and $p_n(x)$ è l n -ma stima di $p(x)$, si ha:

$$p_n(x) = (k_n/n)/V_n$$



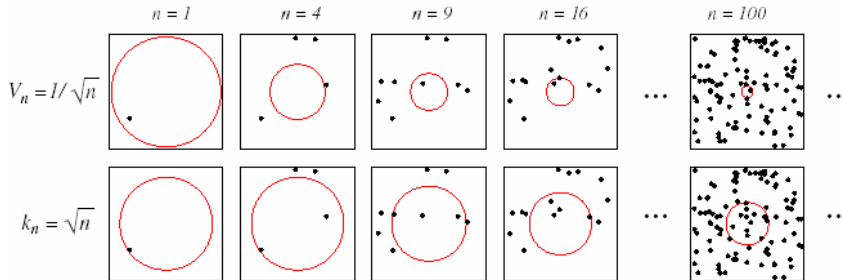
Stima della densità

- Perché $p_n(x)$ converga a $p(x)$ sono necessarie tre condizioni:

$$\lim_{n \rightarrow \infty} V_n = 0 \quad \lim_{n \rightarrow \infty} k_n = \infty \quad \lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$$

- Due modi per ottenere tali condizioni:
 - ridurre la regione R definita inizialmente specificando il volume V_n come funzione di n (es.: $V_n=1/\sqrt{n}$) e dimostrare che $p_n(x) \rightarrow p(x)$ per $n \rightarrow \infty$ (*metodo della finestra di Parzen*).
 - specificare k_n come funzione di n (es.: $k_n = \sqrt{n}$). In questo caso, V_n cresce fino a contenere k_n campioni (*stima a k_n vicini*).

Stima della densità



Stima della densità con i due metodi. Entrambe le sequenze rappresentano variabili aleatorie che generalmente convergono, permettendo di stimare la densità nel punto di interesse.

Metodo della finestra di Parzen



- Assumiamo che la regione R_n sia un ipercubo a d dimensioni, di lato h_n e volume $V_n = h_n^d$.
- Consideriamo una *funzione finestra* $\varphi(u)$ che unitaria all'interno di un ipercubo centrato nell'origine e di lato unitario:

$$\varphi(u) = \begin{cases} 1 & |u_j| \leq \frac{1}{2} \quad j=1, \dots, d \\ 0 & \text{altrimenti} \end{cases}$$

- $\varphi((x-x_i)/h_n)$ è uguale a 1 se x_i cade all'interno dell'ipercubo di volume V_n centrato su x e nullo al di fuori.

Metodo della finestra di Parzen



- Il numero di campioni che cade all'interno di V_n è quindi uguale a:

$$k_n = \sum_{i=1}^n \varphi\left(\frac{x-x_i}{h_n}\right)$$

- La stima della densità è quindi:

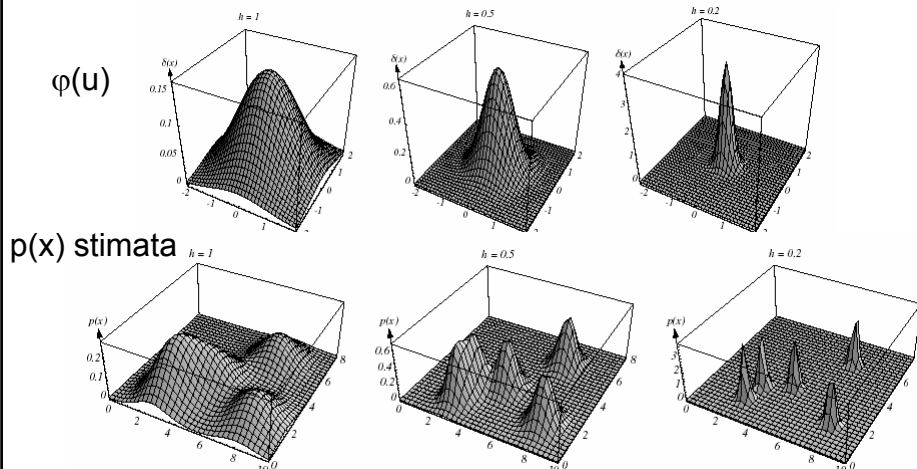
$$p_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \varphi\left(\frac{x-x_i}{h_n}\right)$$

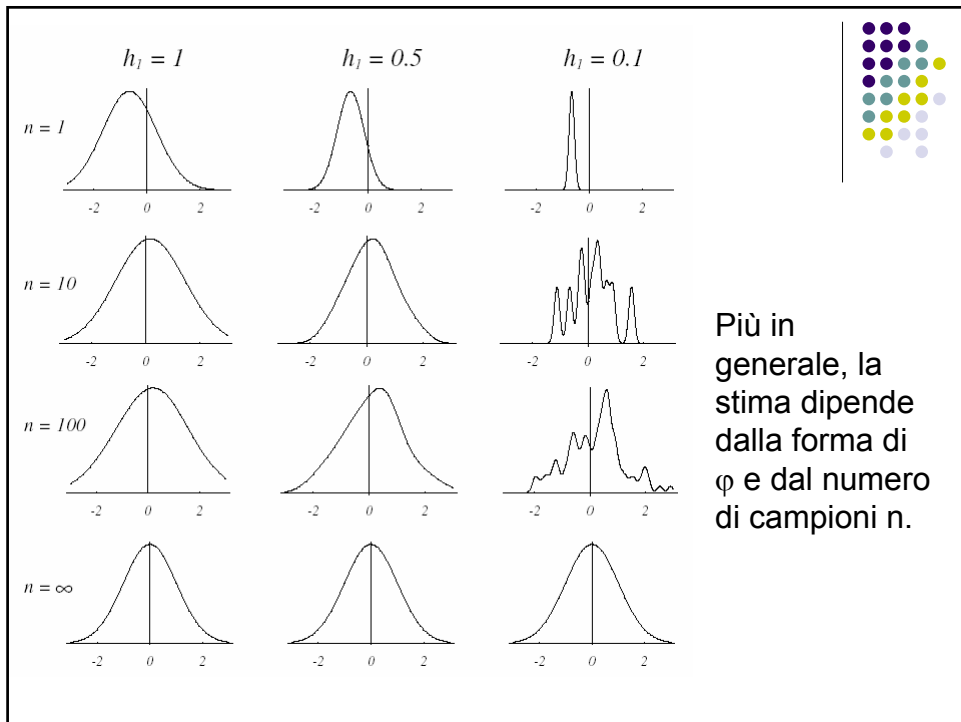
- $p_n(x)$ stima $p(x)$ come la media di funzioni di x e dei campioni (x_i) ($i = 1, \dots, n$). La funzione φ può essere di forma generale purchè si verifichi $\varphi(u) \geq 0$ e $\int \varphi(u) du = 1$.

Metodo della finestra di Parzen



A parità di n , la stima dipende dalla forma della φ :

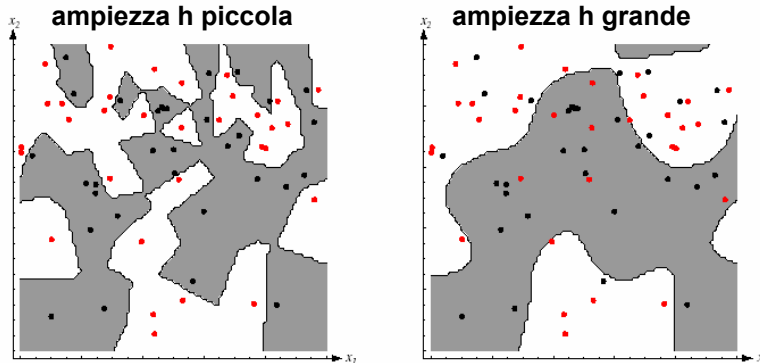




Classificazione

- Nei classificatori basati sulla stima con la finestra di Parzen, si stima la densità condizionata per ogni classe e si esegue la classificazione con la regola della massima probabilità a posteriori.
- Le regioni di decisione dipendono dalla scelta della funzione finestra.

Classificazione



La forma della φ influisce sulle regioni di decisione.

Stima k_n -nearest-neighbor

- Un problema della stima con la finestra di Parzen è la scelta della funzione di finestra.
- Con la stima a k_n vicini (k_n nearest neighbor) si va a stimare direttamente le probabilità a posteriori per ogni classe.
- La classificazione si limita quindi a scegliere la classe con la massima prob. a posteriori .



Stima k_n -nearest-neighbor

- Supponiamo di circondare un punto x appartenente ad un insieme di n campioni con un cella di volume V .
- Se la cella assorbe k campioni, di cui k_i appartenenti alla classe ω_i , una stima della probabilità congiunta $p(x, \omega_i)$ è data da:

$$p_n(x, \omega_i) = \frac{k_i/n}{V}$$

e quindi possiamo ottenere una stima di $P(\omega_i|x)$ come:

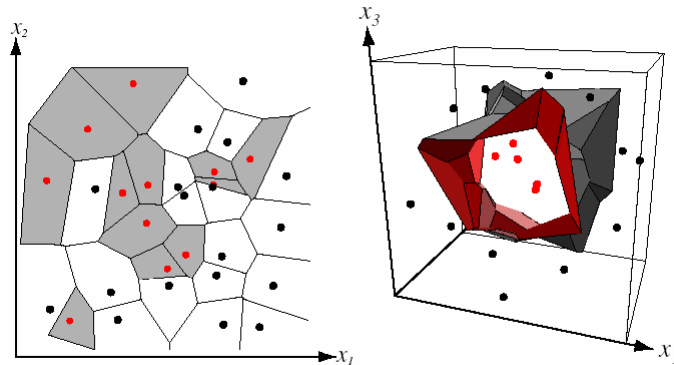
$$P_n(\omega_i | x) = \frac{p_n(x, \omega_i)}{\sum_{j=1}^c p_n(x, \omega_j)} = \frac{k_i}{k}$$



Classificatore Nearest-Neighbor

- E' un classificatore non parametrico che impiega un insieme $D_n = \{x_1, x_2, \dots, x_n\}$ di campioni appartenenti a tutte le classi, detti *prototipi*.
- La classificazione di un nuovo campione x non appartenente a D_n avviene scegliendo l'etichetta del prototipo x' a minima distanza da x .
- Viene denotato con "classificatore NN".

Classificatore Nearest-Neighbor



Il classificatore NN induce sullo spazio delle features una tassellazione di Voronoi.

Classificatore Nearest-Neighbor

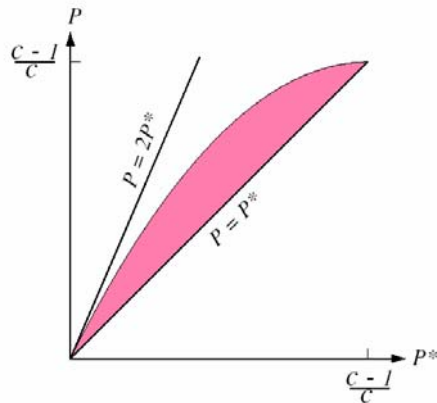


- Il classificatore è sub-ottimo nel senso che non garantisce la probabilità di errore minima esibita dal classificatore bayesiano.
- E' però possibile dimostrare che, al crescere di n , la probabilità di errore P_e per il classificatore NN soddisfa la seguente relazione:

$$P_{e^*} \leq P_e \leq 2P_{e^*}$$

dove P_{e^*} è la probabilità di errore del classificatore bayesiano.

Classificatore Nearest-Neighbor

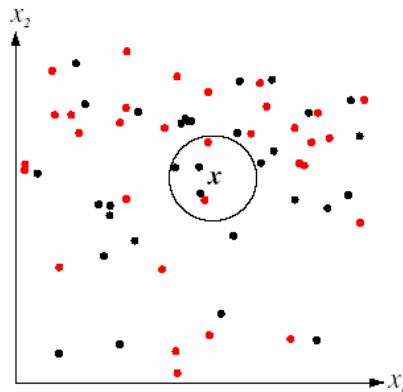


Classificatore k-nearest-neighbor



- Una naturale estensione del classificatore NN è il k-Nearest-Neighbor, denotato con “classificatore k-NN” (perciò il classificatore NN viene spesso denotato con “1-NN”).
- La classificazione di un campione x comporta l’individuazione in D_n dei k punti più vicini a x e la scelta dell’etichetta più rappresentata.
- Per evitare degli ex-aequo tra classi, è necessario scegliere k dispari.

Classificatore k-nearest-neighbor



Classificatore k-nearest-neighbor



All'aumentare di k la probabilità di errore del classificatore k-NN si avvicina alla probabilità del classificatore bayesiano.

