

Riconoscimento e Recupero dell'Informazione per Bioinformatica

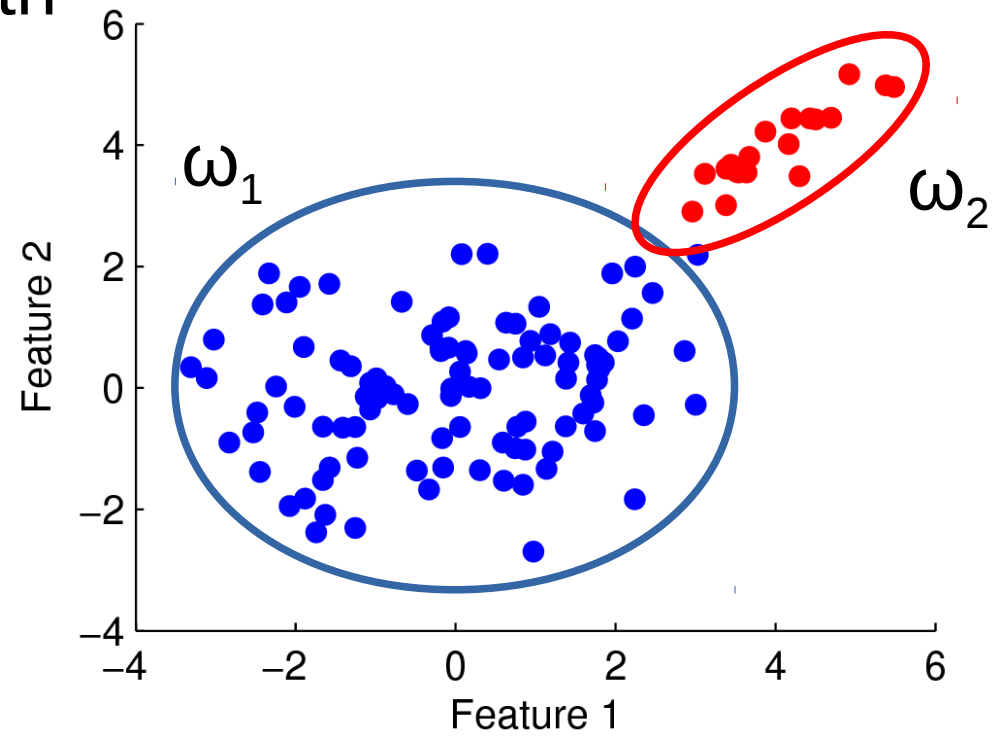
LAB. 5 – Parzen Windows

Pietro Lovato

Corso di Laurea in Bioinformatica
Dip. di Informatica – Università di Verona
A.A. 2016/2017

Ripasso: esercitazione 4

- Stima parametrica:
 - Si crea un modello per ogni classe
 - 1) Assumendo una forma della distribuzione della classe (es. gaussiana)
 - 2) Stimandone i parametri
 - Si classificano i punti controllando la likelihood oppure la posterior



Parzen Windows

- Problema della stima parametrica: si assume che la forma delle densità di probabilità sia nota, ma questa assunzione non può essere fatta in molti problemi di riconoscimento.
- Idea: stimare la pdf andando ad analizzare le singole regioni dello spazio
 - Se bisogna stimare $p(x=x_0)$, si va a considerare la regione attorno ad x_0 e si effettua la stima a partire da quella regione
 - Si può ripetere per tutti i punti da classificare

Come fare

- Punto di partenza:
 - Insieme di dati di train
 - Etichette dei dati di train
 - Insieme di dati di test
 - Etichette dei dati di test (non verranno mai considerate se non nella validazione finale)

```
>> load dataset.mat
```

Come fare

- A priori, decido l'ampiezza della finestra h
 - Per questo esercizio, $h=0.2$
- Divido il dataset (di TRAIN) nelle due classi (dei dati di test non dovrei sapere l'etichetta)

Come fare

Dato un punto di test x_j^{TE} ,

– Per ogni punto di train x_i^{TR} della classe c:

• Calcolo la funzione $\gamma\left(\frac{x_j^{TE} - x_i^{TR}}{h}\right)$

– La likelihood $p(x_j^{TE} | \omega_c) = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{1}{h} \gamma\left(\frac{x_j^{TE} - x_i^{TR}}{h}\right)$

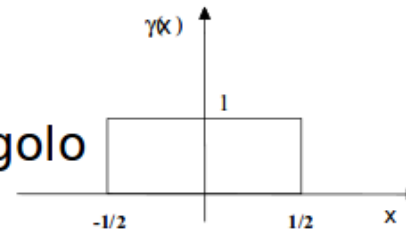
– N_c numero di oggetti nella classe c

Come fare

- Assegno x_j^{TE} alla classe con likelihood massima
- Tipi possibili di funzione γ :

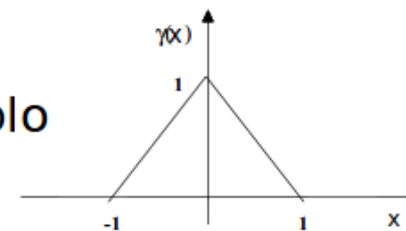
$$1) \gamma(x) = \begin{cases} 1 & |x| \leq 0.5 \\ 0 & |x| > 0.5 \end{cases}$$

Rettangolo



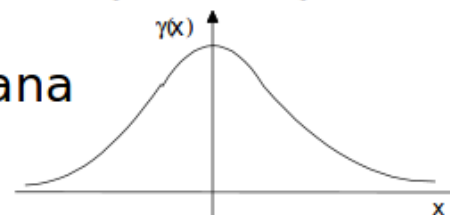
$$2) \gamma(x) = \begin{cases} 1-|x| & |x| \leq 1 \\ 0 & |x| > 1 \end{cases}$$

Triangolo



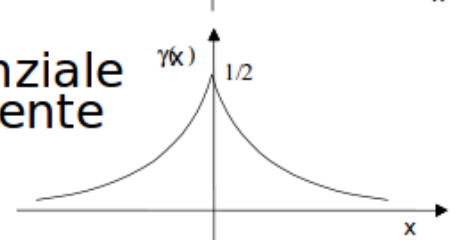
$$3) \gamma(x) = (2\pi)^{-\frac{1}{2}} e^{-\left(\frac{x^2}{2}\right)}$$

Gaussiana



$$4) \gamma(x) = \frac{1}{2} e^{-|x|}$$

Esponenziale
decescente



Esercizio

- Implementare in Matlab un sistema di classificazione attraverso Parzen Windows

