

Riconoscimento e Recupero dell'Informazione per Bioinformatica

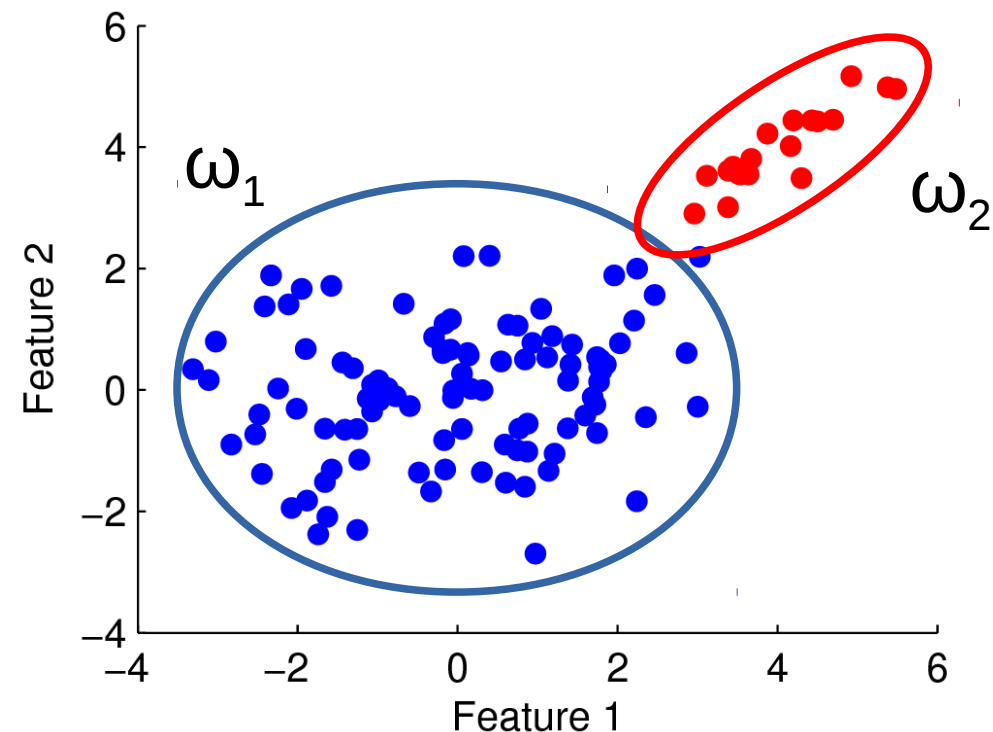
LAB. 4 – Distribuzioni gaussiane

Pietro Lovato

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

Ripasso: Teoria della decisione di Bayes

- Dato un oggetto x , lo assegno alla classe ω_1 oppure a ω_2 ?
- Devo tener conto di:
 - Probabilità a priori
 $P(\omega_1)$
 - Probabilità condizionale
 $p(x|\omega_1)$
 - Probabilità a posteriori
 $p(\omega_1|x)$

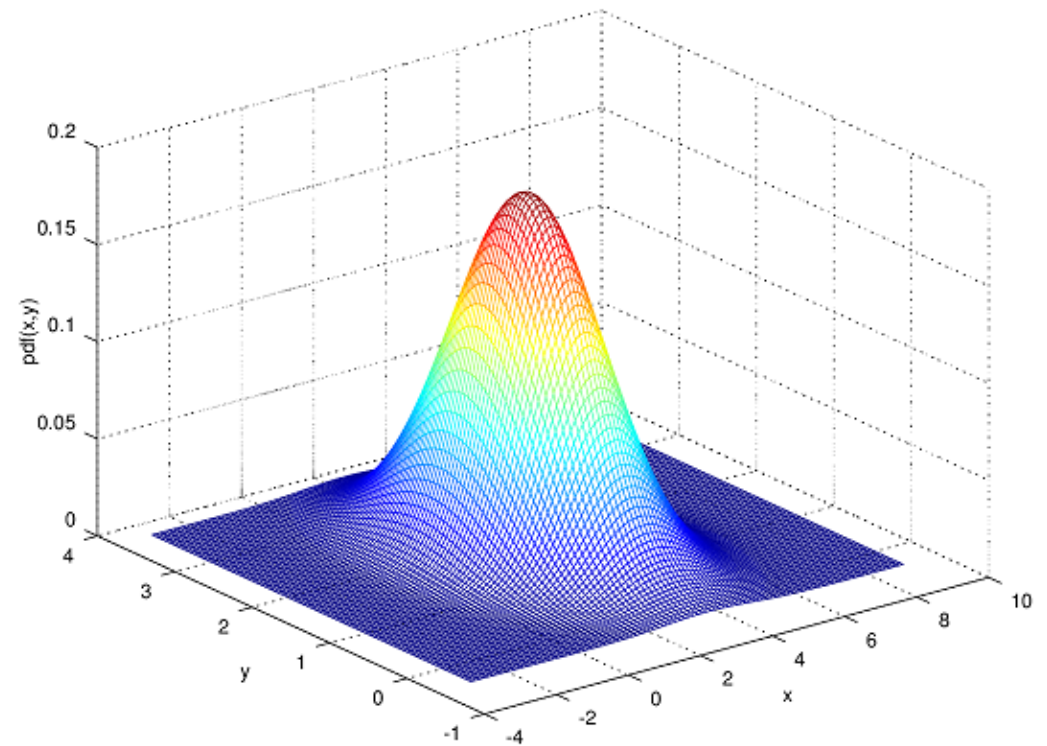
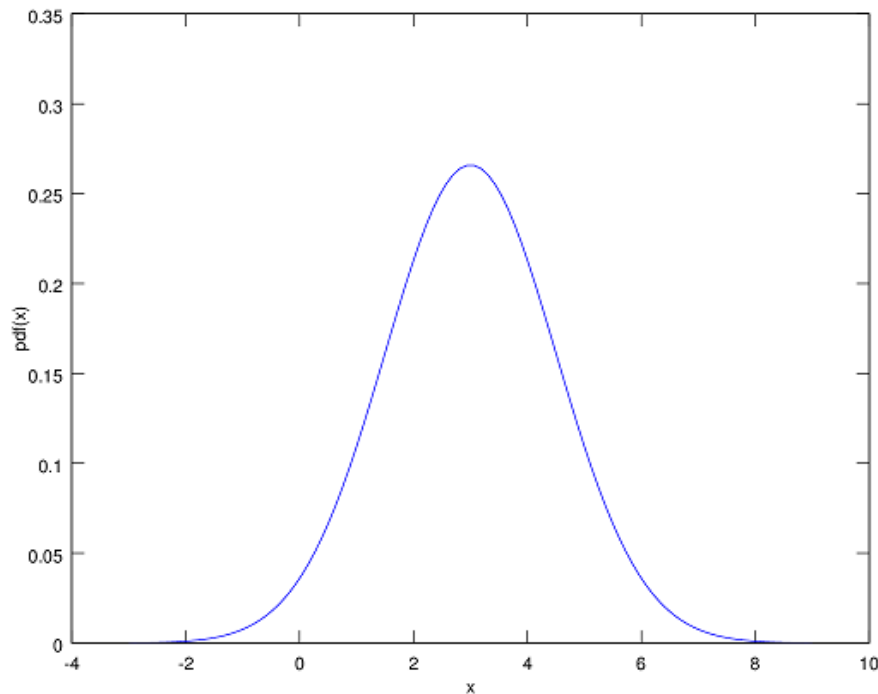


Regola di decisione

- Decidi ω_1 se $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$, ω_2 altrimenti
- Problema: come calcolare le probabilità?

La densità normale

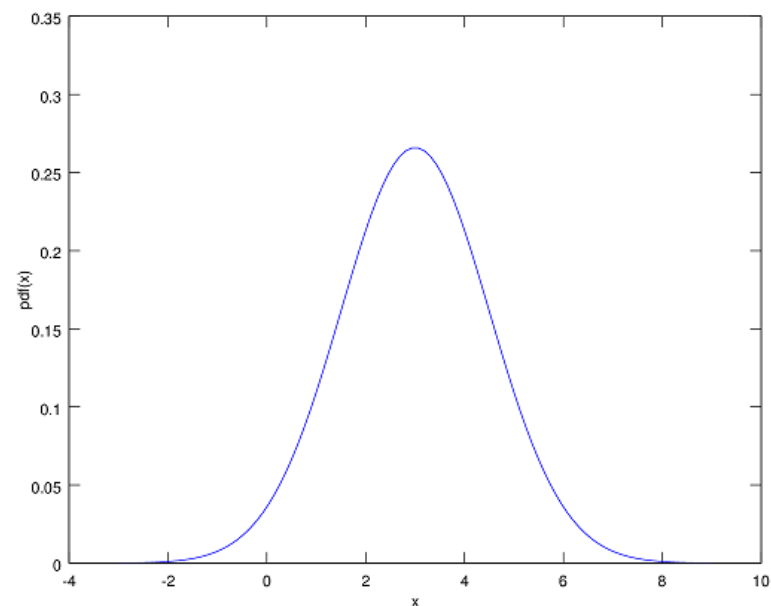
- Assunzione: le probabilità sono gaussiane (multivariate)



La densità normale

- Gaussiana univariata
- Parametri:
 - Media μ
 - Deviazione standard σ

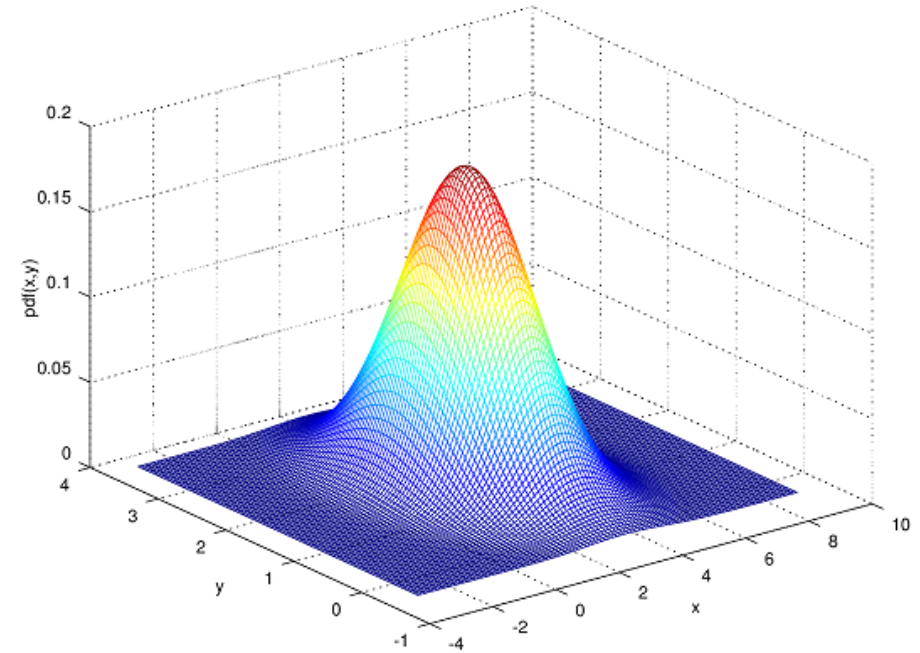
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$



```
>> doc normpdf
```

La densità normale multivariata

- Gaussiana multivariata
- Parametri:
 - Vettore di medie μ
 - Matrice di covarianza Σ



$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right)$$

```
>> doc mvnpdf
```

Esercizio

- Caricare in memoria il file “iris_pca.mat”
- Estrarre 3 diverse sotto-matrici:
 - Gli oggetti della classe 1: sono le colonne di data per cui `labels == 1`
 - Gli oggetti della classe 2
 - Gli oggetti della classe 3
- Stimare media e matrice di covarianza delle 3 classi (suggerimento: per la matrice di covarianza si veda il comando `cov`)

Esercizio

- Valutare la (log)likelihood di ogni oggetto rispetto ai 3 diversi modelli (comando `mvnpdf`)
- Assegnare ogni oggetto alla classe con likelihood maggiore, e calcolare l'errore che si commette confrontando la classe assegnata con l'etichetta “vera” contenuta nel vettore `labels`
- Ripetere l'esercizio introducendo un prior sulle classi nel seguente modo:
 - $P(\omega_1) = 0.1$, $P(\omega_2) = 0.1$, $P(\omega_3) = 0.8$