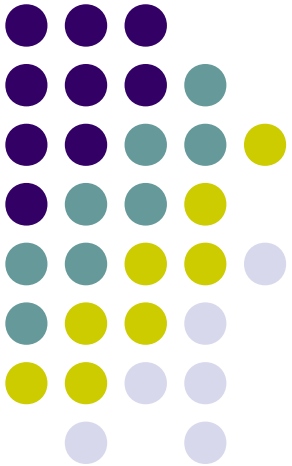
Pattern Recognition

Lab1:
Discriminant Functions

Francesco Tortorella

University of Cassino and Southern Latium

Cassino, Italy



Lab 1.1

- Consider two classes with Gaussian densities:

$$p(x|\omega_1) = N(\mu_1, \Sigma_1)$$
$$p(x|\omega_2) = N(\mu_2, \Sigma_2)$$

- First case $\Sigma_1 = \Sigma_2 = \Sigma$
  - Build two discriminant functions and evaluate the decision boundary for different values of P_2/P_1 {0.1,0.5,1,2,10}

Lab1.1

- Second case $\Sigma_1 \neq \Sigma_2$
  - Build two discriminant functions and evaluate the decision boundary for different values of P_2/P_1 {0.1,0.5,1,2,10}
- Hints:
  - use the Matlab function **`mvnrnd(mu,Sigma,N)`** for plotting the samples of a Gaussian distribution,
  - use the Matlab script `ShowDF` (provided on Classroom) for visualizing decision regions and boundary given two discriminant functions.

Lab 1.2

- Consider the two cases of Lab 1.1, but in a cost-sensitive framework
- Use the following cost-matrices:

$$\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 2 \\ 4 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 4 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 8 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 4 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 8 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} -1 & 4 \\ 1 & -4 \end{pmatrix} \quad \begin{pmatrix} -4 & 4 \\ 1 & -1 \end{pmatrix}$$

Lab 1.3

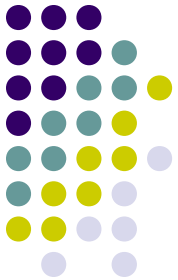
- Class Distribution: (class value 1 is interpreted as "tested positive for diabetes")
  - Class Value Number of instances

    0 500

    1 268
- Brief statistical analysis:

Attribute number:	Mean:	Standard Deviation:
1.	3.8	3.4
2.	120.9	32.0
3.	69.1	19.4
4.	20.5	16.0
5.	79.8	115.2
6.	32.0	7.9
7.	0.5	0.3
8.	33.2	11.8

Lab 1.3



- Read the file ‘pima-indians-diabetes.data’
- Part the set into two subsets: assume one (PimaTr) as training set and the other one (PimaTest) as test set
- Starting from the training set, build a linear classifier and a quadratic classifier and compare the accuracies on the test set
- Consider different sizes of the training set (25%, 50%, 75%) and analyze how the accuracy changes.