ESTIN

Pattern recognition for image analysis

S5(AI-DS) 2024-2025

Lab° 02

Exercise 1:

- Draw the decision boundary between two Gaussian distributions (Case: 1D, 2 classes).

Exercise 2:

Let two coordinates x and y be represented as a two-dimensional vector $\boldsymbol{x}$. We consider two classes ω_1 and ω_2 with data-generating Gaussian distributions $p(\boldsymbol{x}/\omega_1)$ and $p(\boldsymbol{x}/\omega_2)$ of mean vectors $\boldsymbol{\mu}_1 = [-0.5, -0.5], \boldsymbol{\mu}_2 = [0.5, 0.5]$, respectively, and same covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Classes occur with probability $P(\omega_1) = 0.9$ and $P(\omega_2) = 0.1$. Analysis tells us that in such scenario, the optimal decision boundary between the two classes should be linear. We would like to verify this computationally by applying Bayes decision theory on grid-like discretized distributions.

- Discretize the two data-generating distributions $p(\boldsymbol{x}/\omega_1)$ and $p(\boldsymbol{x}/\omega_2)$ and plot them with different colors.
- From these distributions, compute the total probability distribution $p(\boldsymbol{x})$, and plot it.
- Compute and plot the class posterior probabilities $P(\omega_1/x)$ and $P(\omega_2/x)$, and print the Bayes error rate for the discretized case.

Exercise 3:

I.

a) Write a function to calculate **the discriminant function** of the following form

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^t {\Sigma_i}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i) - \frac{d}{2} ln 2\pi - \frac{1}{2} ln|\Sigma_i| + \ln P(\omega_i)$$

for a given mean vector $\boldsymbol{\mu}_i$ and a covariance matrix Σ_i.

b) Write a function to calculate **the Mahalanobis distance** between an arbitrary point $\boldsymbol{x}$ and the mean $\boldsymbol{\mu}_i$, of a Gaussian distribution with covariance matrix Σ.

c) Write a function to calculate **the Minimum distance** between an arbitrary point $\boldsymbol{x}$ and the mean, $\boldsymbol{\mu}_i$.

II. Using functions developed above:

a) $\boldsymbol{\mu}_1 = [3, 6]^t, \boldsymbol{\mu}_2 = [3, -2]^t$. $\Sigma_1 = \begin{bmatrix} 1/2 & 0.3 \\ 0.3 & 2 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. Find the decision boundary and draw it.

b) In a 2 class, 2D, the feature vectors are generated by two normal distribution sharing the same covariance matrix $\Sigma_i = \Sigma = \begin{bmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{bmatrix}$, the mean vectors are $\boldsymbol{\mu}_1 = [0, 0]^t, \boldsymbol{\mu}_2 = [3, 3]^t$. Classify the vector $[1.0, 2.2]^t$ according to Bayesian classifier.

Exercise 4:

- Load the **Iris dataset** from **sklearn**
- Visualize the data using scatter plot (use Petal Length and Petal Width as illustrated in figure below).
- Compute the class means
- If the data is classified using a **minimum distance classifier**, draw the decision boundaries and class means on the plot.