Assignment – 2

Bayes Decision Rule Classifier

Q.1 Generate N = 500 2-dimensional data points that are distributed according to the Gaussian distribution $N(m,S)$, with mean m = [0, 0] ^{T} and covariance matrix S =

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

For the following cases,

$\sigma_1^2 = 0.2, \sigma_2^2 = 2, \sigma_{12} = 0$

$\sigma_1^2 = 2, \sigma_2^2 = 0.2, \sigma_{12} = 0$

$\sigma_1^2 = \sigma_2^2 = 1, \sigma_{12} = 0.5$

$\sigma_1^2 = 0.3, \sigma_2^2 = 2, \sigma_{12} = 0.5$

$\sigma_1^2 = 0.3, \sigma_2^2 = 2, \sigma_{12} = -0.5$

Plot the generated numbers to visualize the distributions.

Q.2 Consider a 2-dimensional classification problem where the data vectors stem from two equiprobable classes, ω_1 and ω_2. The classes are modelled by Gaussian distributions with means m_1 = [0,0] ^{T}, m_2 = [1,2] ^{T}, and respective covariance matrices

$$S_1 = S_2 = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

Generate two data sets X_1 and X_2 consisting of 1000 and 5000 points, respectively. Taking X_1 as the training set, classify the points in X_2 using the squared Euclidean distance-based classifier. Compute the classification error.

Q.3 Generate a set X_1 that consists of N_1 = 50 5-dimensional data vectors that stem from two equiprobable classes, $\omega 1$ and $\omega 2$. The classes are modelled by Gaussian distributions with means $m1$ = [0,0,0,0,0]^T and $m2$ = [1,1,1,1,1] ^{T} and respective covariance matrices

$$S_1 = \begin{bmatrix} 0.8 & 0.2 & 0.1 & 0.05 & 0.01 \\ 0.2 & 0.7 & 0.1 & 0.03 & 0.02 \\ 0.1 & 0.1 & 0.8 & 0.02 & 0.01 \\ 0.05 & 0.03 & 0.02 & 0.9 & 0.01 \\ 0.01 & 0.02 & 0.01 & 0.01 & 0.8 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 0.9 & 0.1 & 0.05 & 0.02 & 0.01 \\ 0.1 & 0.8 & 0.1 & 0.02 & 0.02 \\ 0.05 & 0.1 & 0.7 & 0.02 & 0.01 \\ 0.02 & 0.02 & 0.02 & 0.6 & 0.02 \\ 0.01 & 0.02 & 0.01 & 0.02 & 0.7 \end{bmatrix}$$

In a similar manner, generate a data set X_2 consisting of $N_2 = 10{,}000$ data points. X_1 is used for training; X_2, for testing. In the spirit of the naive Bayes classifier, we assume that for each class the features of the feature vectors are statistically independent and that each follows a 1-dimensional Gaussian distribution. For each of the five dimensions and for each of the two classes, the mean values are m_{1j}, m_{2j}, $j = 1, 2, \ldots, 5$ and the variances are σ^2_{1j}, σ^2_{2j}, $j = 1, 2, \ldots, 5$.

Classify the points of the test set X_2 using the naive Bayes classifier, where for a given x, $p(x|\omega_i)$ is estimated as

$$p(x|\omega_i) = \prod_{j=1}^{5} \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left(-\frac{(x(j) - m_{ij})^2}{2\sigma_{ij}^2}\right), \; i = 1,2$$

where $x(j)$ is the jth component of x. Compute the error probability.

Hint: Use random.gauss (https://docs.python.org/2/library/random.html#random.gauss) function to generate the random points.