Machine Learning Assignment Number 3

Assignments on Artificial Dataset and Real life datasets. Real life dataset availability: UCI machine learning repository; Datasets: Labelled and unlabelled dataset

Artificial Datasets:

$\mu_1' = (0,0)$; $\mu_2' = (0,1)$; $\mu_3' = (1,0)$

$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$; $\Sigma_2 = \begin{pmatrix} 2 & -1 \\ -1 & 1.5 \end{pmatrix}$; $\Sigma_3 = \begin{pmatrix} 2 & 1 \\ 1 & 1.5 \end{pmatrix}$;

Case 1: $P_1 = 0.5 = P_2$

Case 2: $P_1 = 0.4$; $P_2 = 0.6$

A) Consider $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_1)$ under case (i)

Generate n points from mixture density functions. The value of $n =$
100, 500, 1000, 2000, 5000

B) Consider $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_1)$ under case (ii)

Generate n points from mixture density functions randomly.

C) Consider $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ under case (ii)

Generate n points from mixture density functions randomly.

D) Consider $N(\mu_1, \Sigma_1)$, $N(\mu_2, \Sigma_2)$, and $N(\mu_3, \Sigma_3)$

$P_1 = 0.3$, $P_2 = 0.4$, $P_3 = 0.3$
Generate n points from mixture density functions randomly.

1) Take first $\frac{n}{2}$ points as training set and the rest as test set.
2) For each training set and for each class estimate mean, covariance matrix and prior probability.
3) After estimating three, apply Bayes' decision rule on each point in the test set by using he estimated values under the assumption of normality, and find the misclassification rate.
4) Apply Bayes' decision rule on each point of test by using the actual parameter values under the assumption of normality and find the misclassification rate.