In this assignment I have designed and tested a Bayseian, minimum error classifier using Maximum Likelihood Estimation Technique on a set of training data, H, and tested out the performance of the classifier on a set of test data, S_T.

The mathematical formula of the Bayseian density function classifier is of the form:-

$p(x) = \{1/ [(2*\pi)^d * |\Sigma|]^{1/2}\}*\exp[-1/2(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})]$ where Σ represents the estimated covariance of a class and μ represents the estimated mean vector of the classes.

Here are the estimated parameters(mean and covariance)

Estimated Class 1 Mean Vector

$$\begin{matrix} 0.54656947 \\ 50.09870688 \\ -24.68636838 \\ -49.62397174 \end{matrix}$$

Estimated Class 1 Covariance Matrix

$$\begin{bmatrix} 5.14382273e+03 & 6.10661252e+01 & 5.19156287e+01 & 1.38427767e+02 \\ 6.10661252e+01 & 2.43331397e+03 & 3.54336017e+01 & -3.72154574e+01 \\ 5.19156287e+01 & 3.54336017e+01 & 6.23379573e+02 & 4.88330026e+00 \\ 1.38427767e+02 & -3.72154574e+01 & 4.88330026e+00 & 2.59461622e+03 \end{bmatrix}$$

Estimated Class 2 Mean Vector

$$\begin{matrix} -0.08688823 \\ 49.76324894 \\ -25.24143728 \\ 50.54650893 \end{matrix}$$

Estimated Class 2 Covariance Matrix

$$\begin{bmatrix} 2.45481250e+01 & -3.88037864e+00 & -5.17124219e-01 & -1.52078216e+00 \\ -3.88037864e+00 & 2.50730884e+03 & 2.12538596e+01 & 3.94724019e+01 \\ -5.17124219e-01 & 2.12538596e+01 & 6.39553801e+02 & 3.03665035e+01 \\ -1.52078216e+00 & 3.94724019e+01 & 3.03665035e+01 & 5.66340740e+03 \end{bmatrix}$$

Estimated Class 3 Mean Vector

$$\begin{matrix} -5.89344376 \\ -4.59845856 \\ -26.34971881 \\ -49.37342168 \end{matrix}$$

Estimated Class 3 Covariance Matrix

$$\begin{bmatrix} 5786.38115924 & -85.08626397 & 29.45591812 & 89.85211169 \\ -85.08626397 & 2486.55244034 & -62.91748891 & -17.60702012 \\ 29.45591812 & -62.91748891 & 5658.16962549 & -56.56427407 \\ 89.85211169 & -17.60702012 & -56.56427407 & 2457.68384069 \end{bmatrix}$$

Estimated Mean Formula is given by $\mu = 1/n * \sum_{k=1}^{n} x_k$

Estimated Covariance Matrix Formula is given by, $\Sigma_u = 1/(n-1) * \sum_{k=1}^{n}(x - \square)(x - \square)^T$

Since there are 5000 vectors from each class in H, I have assumed the apriori probabilities of each class to be 1/3. Since the apriori probabilities of the classes are equal I have classified each vector in the test set to that class for which the value of $p(x|w_i)$is maximum.

The Confusion Matrix obtained from H is of the following form:-

$$\begin{matrix} 3820 & 479 & 701 \\ 115 & 4817 & 68 \\ 1118 & 181 & 3701 \end{matrix}$$

On the vertical axis we have the true classes and on the horizontal axis we have the classes as classified by the classifier.

The probability of error incurred by the classifier is given by P(Error) = 0.17747