Question 1 (30%)

Design a classifier that achieves minimum probability of error for a three-class problem where the class priors are respectively $p(L=1) = 0.15$, $p(L=2) = 0.35$, $p(L=3) = 0.5$ and the class-conditional data distributions are all Gaussians for two-dimensional data vectors:

$$\mathcal{N}(\begin{bmatrix}-1\\0\end{bmatrix}, [\text{illegible}]) \qquad [\text{illegible}]\mathcal{N}(\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0.5 & 0\\0 & 0.2\end{bmatrix}) \qquad \mathcal{N}(\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}0.1 & 0\\0 & 0.1\end{bmatrix})$$

Generate $N = 10$[illegible] to this data distribution, keeping track of the true class labels for each samp[illegible]mal classifier according to the above probabilities and parameterizations. T[illegible]fier to the dataset and obtain decision labels for each sample. Report the fo[illegible]ng:

- actual number of samples that were generated from each class;
- the confusion matrix for your classifier consisting of number of samples decided as class $r \in \{1,2,3\}$ when their true labels were class $c \in \{1,2,3\}$, using r, c as row/column indices;
- the total number of samples misclassified by your classifier;
- an estimate of the probability of error your classifier achieves based on these samples;
- a visualization of the data as a 2-dimensional scatter plot, with true labels and decision labels denoted using two separate visulization cues, such as marker shape and marker color;
- a clear but brief description of the results presented above.

Question 2 (30%)

Consider a scalar real-valued feature $x \in \mathbb{R}$ that has the following probability distributions under two class labels:

$$p(x|L=1) \; [illegible] \leq 1 \quad [illegible] \leq 2 \qquad\qquad p(x|L=2) = \begin{cases} 2x-3 & \text{if } \frac{3}{2} \leq x \leq \frac{5}{2} \\ 0 & \text{otherwise} \end{cases}$$

Minimum Expec[illegible]cation

Let loss values be [illegible] i when truth is $j) = \lambda_{ij} \geq 0$ for $i, j \in 1, 2$, with loss of erroneous decisions assigned to be greater than corresponding correct decisions. Let the class priors be $q_1 = p(L=1)$ and $q_2 = p(L=2)$, respectively. Express the minimum expected loss decision rule with a discriminant function that is simplified as much as possible. Show your steps.

Maximum a Posteriori Classification

For the case when 0-1 loss (zero-one loss) assignments are used, the minimum expected risk classifier reduces to the maximum a posteriori classification rule. For this case, express the maximum a posteriori classification rule.

Maximum Likelihood Classification

In addition to 0-1 loss assignments, assume that the class priors are equal (uniformly distributed). In this case, minimum expected risk further reduces to maximum likelihood classification. Express the maximum likelihood classification rule for this case.

Question 3 (40%)

Generate N random iid (independent and identically distributed) 2-dimensional samples from two Gaussian pdfs $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with specified prior class probabilities q_i for $i \in 0, 1$. Set

$$\boldsymbol{\mu} \text{ [illegible]} \qquad \boldsymbol{\Sigma}_0 = \begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 16 \end{bmatrix}$$

Make sure to keep tr[illegible] for each sample. To produce numerical results, please use $N = 10000$, $q_0 =$ [illegible] For the following two classification methods, generate classification decisio[illegible]s of their respective thresholds (symbolized as γ). For each classifier, plot t[illegible]sitive and true-positive probabilities for each value of the threshold, i.e. plot their ROC curves. Overlay the two curves in the same plot for easy visual comparison.

Minimum Expected Loss Classifier

Determine and implement the minimum expected loss classifier parameterized by a threshold γ in the following form:

$$\ln p(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) - \ln p(\mathbf{x}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \underset{\text{Decide } 0}{\overset{\text{Decide}}{\gtrless}} \ln \gamma$$

where $\gamma > 0$ is a scalar that depends on the loss values and prior probabilities. Make this threshold take many values along the positive real axis, and for each threshold value, classify every sample. Empirically estimate the true-positive and false-positive probabilities (by counting samples that fall into each category), and then plot the true-positive versus false-positive performance at each threshold using an ROC curve. Sweep the positive real axis by sampling densely to see the true-positive vs false-positive curves for the classifier in detail. *Hint:* sweep through the sorted values for every possible classification split, as determined by the difference in class-conditional log-likelihoods per sample. Here class 1 is *positive* and class 0 is *negative*.

Fisher's Linear Discriminant Analysis (LDA)

Implement the Fisher's LDA classifier using the true mean vectors and covariance matrices provided above to obtain a classifier γ in the following form:

$$\mathbf{w}^{\intercal}\mathbf{x} \underset{\text{Decide } 0}{\overset{\text{Decide } 1}{\gtrless}} \gamma$$

where $\gamma \in (-\inf, \inf)$. Repeat the exercise in the previous section. In other words, first assign many values to this threshold parameter and estimate the true-positive and false-positive probabilities from data classification label matches and mismatches. Then plot the true-positive versus false-positive performance curves as the threshold changes. Overlay this curve on top of the previous one.