EECE5644 2021 Summer10 – Take Home Exam 1

Submit: Monday, 2021-June-01 before 10:00am ET

Please submit your solutions on Canvas in a single PDF file that includes all math, numerical and visual results. Either include a link to your code in an online repository or include the code as an appendix in the PDF file. The code is not graded, but helps verify your results are feasible as claimed. Only results and discussion presented in the PDF will be graded, so do not link to an external location where further results may be presented.

This is a graded assignment and the entirety of your submission must contain only your own work. You may benefit from publicly available literature including software (not from classmates), as long as these sources are properly acknowledged in your submission. All discussions and materials shared during office periods are also acceptable resources and these tend to be very useful, so participate in office periods or take a look at their video recordings. Cite your sources as appropriate.

By submitting a PDF file in response to this take home assignment you are declaring that the contents of your submission, and the associated code is your own work, except as noted in your citations to resources.

Question 1 (50%)

The probability density function (pdf) for a 2-dimensional real-valued random vector $\mathbf{X}$ is as follows: $p(\mathbf{x}) = P(L=0)p(\mathbf{x}|L=0) + P(L=1)p(\mathbf{x}|L=1)$. Here L is the true class label that indicates which class-label-conditioned pdf generates the data.

The class priors are $P(L=0) = 0.65$ and $P(L=1) = 0.35$. The class class-conditional pdfs are $p(\mathbf{x}|L=0) = w_1 g(\mathbf{x}|\mathbf{m}_{01}, \mathbf{C}_{01}) + w_2 g(\mathbf{x}|\mathbf{m}_{02}, \mathbf{C}_{02})$ and $p(\mathbf{x}|L=1) = g(\mathbf{x}|\mathbf{m}_1, \mathbf{C}_1)$, where $g(\mathbf{x}|\mathbf{m}, \mathbf{C})$ is a multivariate Gaussian probability density function with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{C}$. The parameters of the class-conditional Gaussian pdfs are: $w_1 = w_2 = 1/2$, and

$$\mathbf{m}_{01} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \mathbf{C}_{01} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{m}_{02} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \quad \mathbf{C}_{02} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad \mathbf{m}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \mathbf{C}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

For numerical results requested below, generate 10000 samples according to this data distribution, keep track of the true class labels for each sample. Save the data and use the same data set in all cases.

Part A: ERM classification using the knowledge of true data pdf:

1. Specify the minimum expected risk classification rule in the form of a likelihood-ratio test: $\frac{p(\mathbf{x}|L=1)}{p(\mathbf{x}|L=0)} \overset{?}{>} \gamma$, where the threshold γ is a function of class priors and fixed (nonnegative) loss values for each of the four cases $D = i|L = j$ where $D \in 0, 1$ is the decision.

2. Implement this classifier and apply it on the 10K samples you generated. Vary the threshold γ gradually from 0 to ∞, and for each value of the threshold estimate $P(D=1|L=1;\gamma)$ and $P(D=1|L=0;\gamma)$. Using these paired values, plot an approximation of the ROC curve of the minimum expected risk classifier. Note that at $\gamma = 0$ The ROC curve should be at $\binom{1}{1}$, and as γ increases it should traverse towards $\binom{0}{0}$. Due to the finite number of samples used to estimate probabilities, your ROC curve approximation should reach this destination value for a finite threshold value.

3. Determine the theoretically optimal threshold value that achieves minimum probability of error, and on the ROC curve, superimpose (using a different color/shape marker) the operating point of this min-P(error) decision rule by evaluating its two confusion matrix entries needed for its ROC curve point. Estimate the minimum probability of error that is achievable for this data distribution using the dataset and this theoretically optimal threshold. In addition, by evaluating estimated $P(error;\gamma) = P(D=1|L=0;\gamma)P(L=0) + P(D=0|L=1;\gamma)P(L=1)$ values, determine empirically using the dataset a threshold value that minimizes this estimated P(error) value. How does your empirically determined γ value that minimizes P(error) compare with the theoretically optimal threshold you compute from priors and loss values?

Part B: Repeat the same steps as in the previous two cases (draw ROC curve & find threshold that minimizes P(error)), but this time using a Fisher Linear Discriminant Analysis (LDA) based classifier. Using the 10K available samples and their labels, estimate the class conditional mean and covariance matrices using sample average estimators. From these estimates, determine the Fisher LDA projection weight vector (via the generalized eigendecomposition of within and between class scatter matrices): $\mathbf{w}_{LDA}$. For the classification rule $\mathbf{w}_{LDA}^T \mathbf{x}$ compared to a threshold τ, which

takes values from $-\infty$ to ∞, plot the ROC curve. Identify the threshold at which the probability of error (based on sample count estimates) is minimized, and mark that operating point on the ROC curve estimate. Discuss how this LDA classifier performs relative to the previous two classifiers. Compare the P(error) achieved by LDA with that of the optimal design.

Note: When finding the Fisher LDA projection matrix, do not be concerned about the difference in the class priors. When determining the between-class and within-class scatter matrices, use equal weights for the class means and covariances, like we did in class. You could argue for a weighted approach, but in this case the model mismatch is a more serious issue to be concerned about.

Question 2 (50%)

A 3-dimensional random vector $\mathbf{X}$ takes values from a mixture of four Gaussians. One of these Gaussians represent the class-conditional pdf for class 1, and another Gaussian represents the class-conditional pdf for class 2. Class 3 data originates from a mixture of the remaining 2 Gaussian components with *equal weights*. For this setting where labels $L \in \{1,2,3\}$, pick your own class-conditional pdfs $p(\mathbf{x}|L=j),\ j \in \{1,2,3\}$ as described. Try to approximately set the distances between means of pairs of Gaussians to twice the average standard deviation of the Gaussian compoenents, so that there is some significant overlap between class-conditional pdfs. Set class priors to $0.3, 0.3, 0.4$.

Part A: Minimum probability of error classification (0-1 loss, also referred to as Bayes Decision rule or MAP classifier).

1. Generate 10000 samples from this data distribution and keep track of the true labels of each sample.
2. Specify the decision rule that achieves minimum probability of error (i.e., use 0-1 loss), implement this classifier with the true data distribution knowledge, classify the 10K samples and count the samples corresponding to each decision-label pair to empirically estimate the confusion matrix whose entries are $P(D=i|L=j)$ for $i,j \in \{1,2,3\}$.
3. Provide a visualization of the data (scatter-plot in 3-dimensional space), and for each sample indicate the true class label with a different marker shape (dot, circle, triangle, square) and whether it was correctly (green) or incorrectly (red) classified with a different marker color as indicated in parentheses.

Part B: Repeat the exercise for the ERM classification rule with the following loss matrices which respectively care 10 times or 100 times more about not making mistakes when $L=3$:

$$\Lambda_{10} = \begin{bmatrix} 0 & 1 & 10 \\ 1 & 0 & 10 \\ 1 & 1 & 0 \end{bmatrix} \quad \textit{and} \quad \Lambda_{100} = \begin{bmatrix} 0 & 1 & 100 \\ 1 & 0 & 100 \\ 1 & 1 & 0 \end{bmatrix} \tag{1}$$

Note that, the $(i,j)^{th}$ entry of the loss matrix indicates the loss incurred by deciding on class i when the true label is j. For this part, using the 10K samples, estimate the minimum expected risk that this optimal ERM classification rule will achieve. Present your results with visual and numerical reprentations. Briefly discuss interesting insights, if any.