

EECE5644 Spring 2020 – Take Home Exam 1

Submit: Monday, 2020-February-10 before 11:45ET

Please submit your solutions on Blackboard in a single PDF file that includes all math and numerical results. Only the contents of this PDF will be graded.

For control purposes, please also provide any code you write to solve the questions in one of the following ways:

1. Include all of your code in the PDF as an appendix.
2. In the PDF provide a link to your online code repo where the code resides (e.g. GitHub, Bitbucket, GitLab). Do NOT make your repo public until after the submission deadline. If people find your code and use it to solve questions, we will consider it cheating for all parties involved.

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, and as allowed by specific restrictions in questions, if any), as long as these sources are properly acknowledged in your submission. Copying math or code from each other is clearly not allowed and will be considered in the framework of academic dishonesty. Discussing with instructor and teaching assistant to get clarification or to eliminate doubts is acceptable. 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 acknowledgement citations to resources.

Start early and use the office periods well. The office periods are on gcal and contain hangout links for remote participants to our class. You may call 1-617-3733021 to reach Deniz to request a video meeting at the office periods.

Question 1 (60%)

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 $P(L = 0) = 0.8$ and $P(L = 1) = 0.2$. The class class-conditional pdfs are $p(\mathbf{x}|L = 0) = g(\mathbf{x}|\mathbf{m}_0, \mathbf{C}_0)$ 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:

$$\mathbf{m}_0 = \begin{bmatrix} -0.1 \\ 0 \end{bmatrix} \quad \mathbf{C}_0 = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix} \quad \mathbf{m}_1 = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix} \quad \mathbf{C}_1 = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

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.

Minimum expected risk classification using the knowledge of true data pdf:

1. Specify the minimum expected risk classification rule in the form of a likelihood-ratio test: $\frac{p(\mathbf{x}|L=1)}{p(\mathbf{x}|L=0)} \stackrel{?}{>} \gamma$, 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 is the decision label that is either 0 or 1, like L .
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 compute the true positive (detection) probability $P(D = 1|L = 1)$ and the false positive (false alarm) probability $P(D = 1|L = 0)$. Using these values, trace/plot an approximation of the ROC curve of the minimum expected risk classifier.
3. Determine the threshold value that achieves minimum probability of error, and on the ROC curve, superimpose clearly (using a different color/shape marker) the true positive and false positive values attained by this minimum-P(error) classifier. Calculate and report an estimate of the minimum probability of error that is achievable for this data distribution.

Classification with incorrect knowledge of the data distribution (Naive Bayesian Classifier, which assumes features are independent given each class label): For the following items, assume that you know the true class prior probabilities and that you think the class conditional pdfs are both Gaussian with the true means, but (incorrectly) with covariance matrices both equal to the identity matrix. Analyze the impact of this model mismatch in this Naive Bayesian (NB) approach to classifier design.

1. Specify the minimum expected risk classification rule in the form of a likelihood-ratio test: $\frac{p_{NB}(\mathbf{x}|L=1)}{p_{NB}(\mathbf{x}|L=0)} \stackrel{?}{>} \gamma$, where the class conditional pdfs are incorrectly known as specified in the naive Bayesian approximation above.
2. Implement this naive-Bayesian 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 compute the true positive (detection) probability $P(D = 1|L = 1)$ and the false positive (false alarm) probability $P(D = 1|L = 0)$. Using these values, trace/plot an approximation of the ROC curve of the minimum expected risk decision rule.

3. Determine the threshold value that achieves minimum probability of error, and on the ROC curve, superimpose clearly (using a different color/shape marker) the true positive and false positive values attained by this naive-Bayesian model based minimum-P(error) classifier. Calculate and report an estimate of the minimum probability of error that is achievable by the naive-Bayesian classification rule for this (true) data distribution.

In the third part of this exercise, repeat the same steps as in the previous two cases for the Fisher Linear Discriminant Analysis based classifier. Using the 10000 available samples, with sample average based estimates for mean and covariance matrix for each class, determine the Fisher LDA projection vector \mathbf{w}_{LDA} . For the classification rule $\mathbf{w}_{LDA}^T \mathbf{x}$ compared to a threshold τ , which takes values from $-\infty$ to ∞ , trace the ROC curve, identify the threshold at which the probability of error (based on sample count estimates) is minimized, and clearly mark that operating point on the ROC curve.

Note: In order for us to have a uniform solution across all submissions, 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.

Question 2 (30%)

For class labels 0 and 1, pick two class priors and two class-conditional pdfs (both in the form of mixtures of two Gaussians). Do not set the class priors to be equal. Use four different Gaussian components when constructing the class-conditional pdfs. Within each class-conditional pdf, do not select the mixture coefficients to be equal. Select your Gaussian mixtures to create an interesting/challenging example.

1. Provide scatter plots of 1000 total samples from this data distribution. Note that both the true class label and within that class, which Gaussian component generates each sample should be selected randomly, in accordance with class prior probabilities, and Gaussian component probabilities (weights), respectively. Do NOT specify number of samples for any of the labels or components. Indicate true class label for each sample with a different marker shape in the scatter plot.
2. Determine the minimum-P(error) classification rule, specify it, draw its decision boundary superimposed on the scatter plot of data, and classify each sample with this classifier and with color cues indicate if the samples are correctly or incorrectly classified. Using your data samples, calculate an estimate of the smallest probability of error achievable for this dataset and report it.

Question 3 (10%)

For a scalar random variable that may be generated by one of two classes, where class priors are equal and class-conditional pdfs are both unit-variance Gaussians with mean values -2 and 2, determine the classification rule that achieves minimum probability of error. Also express the smallest error probability achievable by this classifier in terms of definite integrals of Gaussian class conditional pdfs involved.