Assignment 3

1. **Optimal threshold on classification probability** (pt= 10)

Consider a case where we have learned a conditional probability distribution P(y|x). Suppose there are only two classes, and let $p_0 = P(Y = 0 | x)$ and $p_1 = P(Y = 1 | x)$. Consider the loss matrix below:

Predicted label $\hat{y}$	**True label y**	
	0	1
0	0	λ_{01}
1	λ_{10}	0

a. Show that the decision $\hat{y}$ that minimizes the expected loss is equivalent to setting a probability threshold θ and predicting $\hat{y} = 0$ if $p_1 < \theta$ and $\hat{y} = 1$ if $p_1 \geq \theta$. What is θ as a function of λ_{01} and λ_{10}? (Show your work.)

b. Show a loss matrix where the threshold is 0.1. (Show your work.)

2. **Reject option in classifiers** (pt=15)

n many classification problems one has the option either of assigning x to class j or, if you are too uncertain, of choosing the reject option. If the cost for rejects is less than the cost of falsely classifying the object, it may be the optimal action. Let αi mean you choose action i, for i =1:C +1, where C is the number of classes and C +1is the reject action. Let Y = j be the true (but unknown) state of nature. Define the loss function as follows

$$\lambda(\alpha_i | Y = j) = \begin{cases} 0 & if\ i = j\ and\ i, j \in \{1, \cdots, C\} \\ \lambda_r & if\ i = C + 1 \\ \lambda_s & otherwise \end{cases}$$

In other words, you incur 0 loss if you correctly classify, you incur λ_r loss (cost) if you choose the reject option, and you incur λ_s loss (cost) if you make a substitution error (misclassification).

a. Show that the minimum risk is obtained if we decide $Y = j$ if $p(Y = j|x) \geq p(Y = k|x)$ for all k (i.e., j is the most probable class) and if $p(Y = j|x) \geq 1 - \lambda_r/\lambda_s$; otherwise, we decide to reject. b. Describe qualitatively what happens as λr/λs is increased from 0 to 1 (i.e., the relative cost of rejection increases).

b. Describe qualitatively what happens as λ_r/λ_s is increased from 0 to 1 (i.e., the relative cost of rejection increases).

c. Let's assume we have a two-class classifier problem. These two classes have an equal prior ($\pi_1 = \pi_2$), and the conditional distributions of the feature x given the class labels are defined by $P(x|Y=1) \sim N(1,1)$ and $P(x|Y=2) \sim N(-1,1)$. Find the decision boundaries for

I. λ_r/λ_s=0

II. λ_r/λ_s=1/2

III. λ_r/λ_s=1

3. Reject option in classifiers (pt=5)

Let B = p(D|H1)/p(D|H0) be the Bayes factor in favor of model 1. Suppose we plot two ROC curves, one computed by thresholding B, and the other computed by thresholding p(H1|D). Will they be the same or different? Explain why.

4. Estimation of σ^2 when μ is known (pt=5)

Suppose we sample $x_1, \ldots, x_N \sim N(\mu, \sigma^2)$ where μ is a known constant. Derive an expression for the MLE for σ^2 in this case. Is it unbiased?

5. Parametric and Empirical ROC curve (pt = 35)

We have a dataset with 3000 data points (**data_3_5.csv – first column is the x, the second column is y**). We have checked the data and it suggests that there are two classes of the data. We are interested in building the ROC curve to find an optimal decision.

a. Let's assume the feature per each class has a normal distribution. Build elements of a Bayes classifier and discuss the model parameters.

b. Discuss the fit between the data and your model estimate

c. Build your ROC curve using the Bayes classifier you built in the step a and discuss the result

d. Build your ROC curve empirically

e. Discuss if you suggest any other decision model to reach a better classification performance

f. Model selection: for class 1, we can model it using the following conditional distribution function

$$P(x|y=1) \sim \frac{3}{4}\,N(x;-3,1) + \frac{1}{4}\,N(x;7,0.316)$$

find the likelihood of the data in the class one given this model and what you built in the part a.

g. BIC and Bayes factor: Find the likelihood ratio of two models we have for the data in class 1 (step a and f) and argue which model you pick for the data. Calculate the BIC for these two models – for the second model, we have 5 free parameters.

6. **Naïve Bayes, QDA, and LDA classifier** (pt = 30)

We have a three-class classifier problem (**data_3_6.csv**). We are interested in comparing three different classifiers' performance on this dataset. Consider 5-fold cross-validation to build your classifiers and also in your performance analysis.

a. Visualize the data and discuss how you build different classifier models
b. What are the classes' priors
c. Build a Naïve Bayes classifier. Provide the confusion matrix and report the performance accuracy. Show the decision boundary for your data.
d. Build a QDA classifier. Provide the confusion matrix and report the performance accuracy. Show the decision boundary for your data.
e. Build an LDA classifier. Provide the confusion matrix and report the performance accuracy. Show the decision boundary for your data.
f. Discuss different classifiers' performance and whether you want to consider other classifiers that might have better performance accuracy.