ECE 2372 - Homework 1

Due: 2/2/22

1. Consider n independent tosses of m biased coins. The probability of heads for each coin is p. We know that the exact probability of k heads in n tosses is calculated by the binomial distribution, and we can find an empirical estimate by finding the ratio of the number of times the coin lands on heads to n.

$$\hat{p_i} = \frac{\text{number of times coin} i \text{lands on heads}}{n} \quad \text{where} \quad i = 1, ..., m$$

   a. Fill the table with the exact probability values that at least one coin out of m will have no heads when you flip each coin $n = 10$ times.

	$m = 1$	$m = 1,000$	$m = 1,000,000$
$p = 0.04$			
$p = 0.075$			

   b. In this part, we will compare the the exact probability with its bound suggested by Hoeffing's inequality. Consider ten tosses of two coins with $p = 0.5$, calculate exact probability of the event stated below.

Worst probability?
What is this asking?

$$\mathbb{P}[max_i \mid \hat{p_i} - p_i \mid > \epsilon]$$

   Sketch this probability as a function of $\epsilon \in [0, 1]$. On the same figure, show the Hoeffding's bound. Comment on the tightness of the bound.

2. For X that is a discrete random variable with a probability mass function $g_k(x)$, prove that the expression given below minimizes the probability of error for a classifier with k classes. Note that we have examine this for the continuous case in Lecture 3.

$$\begin{aligned} f^*(x) &= \arg\max_k \mu_k(x) \\ &= \arg\max_k \pi_k \; g_k(x) \end{aligned}$$

3. Suppose we have a scalar input x and we are considering a binary classification problem with the following class conditional probabilities

$$\begin{aligned} X|Y = 0 &\sim g_0(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}\ |x|} \quad \text{and} \\ X|Y = 1 &\sim g_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \end{aligned}$$

a. Sketch $g_0(x)$ and $g_1(x)$ overlayed in one figure.

b. Let the class prior probabilities be equal, e.g., $\pi_0 = \pi_1$. Express the optimum classification rule in terms of minimizing the risk or the probability of error). How would you relate this rule to your sketch of class conditional probabilities in part (a)?

c. Calculate the Bayes risk for the classification rule you derived in part (b).

4. In today's lecture (2/1) we have discussed LDA and obtained a linear classifier as below (see Lecture 4.pdf)

$$\hat{f}(x) = \begin{cases} 1 & if\ a^T x + b \geq 0 \\ 0 & if\ o.w. \end{cases}$$

Please provide a derivation for a and b in this decision rule.

5. For this problem I would like you to implement LDA and a slight variation of it. Then you will test on some data provided.

a. Implement LDA in MATLAB or Python using the lecture 4 notes. Please come up with your version of LDA as it is simple enough to implement.

b. Now implement another version of your algorithm by letting $\Sigma = \sigma^2 I$, here you can use the same covariance estimate you used in part(a) for LDA and replace it with $\frac{1}{d} trace(\hat{\Sigma}) I$.

c. Run your algorithms on the synthetic data sets. You will find four different synthetic data sets posted on Canvas, and compare and commend on the performance of both algorithms on all of these sets.

d. Apply both algorithms on "trainLDA.mat" to train your LDA classifier and test it on "testLDA.mat". Show the testing error you get with both of your classifiers. Do not use the test data in training.

Please submit all of your code for this example so I can recreate your results. You can zip everything into one folder or use even nicer use Jupyter or Matlab Live editor so my work in grading gets easier.