CSE575 HW01, Monday, 09/25/2023, Due: Friday, 10/06/2023

Please note that you have to typeset your assignment using either LATEX or Microsoft Word, and produce a PDF file for submission. Hand-written assignment (or photo of it) will not be graded. You need to submit an electronic version (in PDF form) on the canvas. You should name your file using the format CSE575-HW01-LastName-FirstName.

1. **Probability, MLE, and PAC [20 pts: 4+4+4+4+4]**

   (A) Suppose that X and Y are **independent** events, and $p(Y) > 0, p(X) = 0.5$. What is the value of $p(X|Y)$?

   (B) Suppose that X and Y are **disjoint** events (i.e. $p(X, Y) = 0$) and $p(Y) > 0$. What is the value of $p(X|Y)$?

   (C) Suppose that we have two coins C_1 and C_2. The probability of C_1 having head is 0.6, and the probability of C_2 having head is 0.4. In each test, we toss both coins, and read the faces of C_1 and C_2 (note that we read C_2 **after** reading C_1). For example, if the toss resulted in C_1 head up and C_2 tail up, we will record the result as HT. Suppose we perform the test 4 times. What is the probability for us to observe the following result?

   $$HT,\ HT,\ TT,\ TT?$$

   (D) You are given a coin and are asked to toss as many times as you wish to decide the probability of having heads-up for a toss of the coin. You tossed the coin 20 times, and observed 15 heads and 5 tails. What is your best estimate of the probability θ of having heads-up?

   (E) If you want to be at least 99% sure that the difference between your estimated value of θ and the true probability of the coin having heads-up is no more than 0.1, how many tosses can guarantee this (hint: use the Hoeffding's inequality on slide 11 of Lecture05)? Please give the minimum number of tosses.

2. Discriminant Linear Classifiers [20 pts: 10+10]

You are given a training data set $\{x_n, t_n\}$ of size $N = 21$. Each input vector x_n is a point in the 2-dimensional Euclidean space R^2. We have $x_1 = (0,0)^{\mathrm{T}}$, $x_2 = (1,0)^{\mathrm{T}}$, $x_3 = (2,0)^{\mathrm{T}}$, $x_4 = (0,1)^{\mathrm{T}}$, $x_5 = (1,1)^{\mathrm{T}}$, $x_6 = (2,1)^{\mathrm{T}}$, $x_7 = (3,1)^{\mathrm{T}}$, $x_8 = (4,1)^{\mathrm{T}}$, $x_9 = (5,1)^{\mathrm{T}}$, $x_{10} = (100,1)^{\mathrm{T}}$, $x_{11} = (0,2)^{\mathrm{T}}$, $x_{12} = (1,2)^{\mathrm{T}}$, $x_{13} = (2,2)^{\mathrm{T}}$, $x_{14} = (3,2)^{\mathrm{T}}$, $x_{15} = (4,2)^{\mathrm{T}}$, $x_{16} = (5,2)^{\mathrm{T}}$, $x_{17} = (100,2)^{\mathrm{T}}$, $x_{18} = (3,3)^{\mathrm{T}}$, $x_{19} = (4,3)^{\mathrm{T}}$, $x_{20} = (5,3)^{\mathrm{T}}$, and $x_{21} = (100,3)^{\mathrm{T}}$. Each point is represented as a column vector.

There are two target classes C_1 and C_2. For each point x_n in the training set, x_n belongs to C_1 if its second coordinate is less than or equal to 2, and belongs to C_2 otherwise.

(A) Compute the least-square linear classifier based on the training data (using $K = 2$ in slides of Lecture08 or textbook chapter 4.1.3). You need to write out (a) the error function [5pts], (b) the computed parameter matrix $\widetilde{\mathrm{W}}$ (a 3 by 2 matrix) [5pts].

(B) Compute the linear classifier based on the training data using Fisher's linear discriminant by $\mathbf{w} = \mathbf{S}_{\mathrm{W}}^{-1}(\boldsymbol{m}_2 - \boldsymbol{m}_1)$ where $\mathbf{S}_{\mathrm{W}}$ is the within-class covariance matrix. You need to write out (a) the Fisher criterion [5pts], (b) the computed parameter $\mathbf{w} = (w_0, w_1)^{\mathrm{T}}$ [5 pts].

3. Continuous Bayes Classifier [20 pts: 5+5+5+5]

We want to build a Bayes classifier for a binary classification task ($y = 1$ or $y = 2$) with a 1-dimensional input feature (x). We know the following quantities: (1) $p\,(y = 1) = 0.6$; (2) $p\,(x|y = 1) = 0.5$ for $0 \le x \le 2$ and $p\,(x|y = 1) = 0$ otherwise; and (3) $p\,(x|y = 2) = 0.125$ for $0 \le x \le 8$ and $p\,(x|y = 2) = 0$ otherwise.

(A) What is the prior for class label $y = 2$?

(B) What is $p\,(y = 1|x)$ for $0 \le x \le 8$?

(C) For $x = 1$, what is the class label your classifier will assign? Why? What is the risk of this decision?

(D) What are the decision regions of your Bayes classifier?

4. Discrete Bayes Classifier [20 pts: 5+5+5+5]

We want to build a Bayes classifier for a binary classification task ($y = 1$ or $y = 2$) with input feature x of two binary features (x_1 and x_2). We know the following quantities: (1)

$p(y = 1) = 0.6$; (2) $p(x_1 = 0, x_2 = 0|y = 1) = 0.3$, $p(x_1 = 0, x_2 = 1|y = 1) = 0.1$, $p(x_1 = 1, x_2 = 0|y = 1) = 0.4$, $p(x_1 = 1, x_2 = 1|y = 1) = 0.2$, and (3) $p(x_1 = 0, x_2 = 0|y = 2) = 0.4$, $p(x_1 = 0, x_2 = 1|y = 2) = 0.3$,$p(x_1 = 1, x_2 = 0|y = 2) = 0.2$, $p(x_1 = 1, x_2 = 1|y = 2) = 0.1$.

(A) What is the prior for class label $y = 2$?

(B) What is $p\ (y = 1|x)$?

(C) For an example with $x_1 = 0$ and $x_2 = 1$, what is the class label your classifier will assign? Why? What is the risk of this decision?

(D) What are the decision regions of your Bayes classifier?

5. Naive Bayes Classifier [20 pts:5+10+5]

Given the training data set in Table 2, we want to train a binary classifier using Naive Bayes, with (1) the last column being the class label y , and (2) each column of X being a binary feature.

Input Feature $X = (x_1, x_2, x_3, x_4, x_5)$					Class Label y
Sky	**Temp**	**Humid**	**Wind**	**Water**	**Enjoy Sport**
sunny	warm	normal	strong	warm	Yes
rainy	cold	high	mild	warm	No
sunny	warm	high	mild	warm	Yes
rainy	cold	high	strong	warm	No
sunny	warm	high	strong	cool	Yes
sunny	cold	normal	mild	warm	Yes
rainy	cold	normal	mild	cool	No

Table 2: Training Data Set for Naive Bayes Classifier

(A) How many independent parameters are there in your Naive Bayes classifier? What are they (only list the independent parameters)? Justify your answer.

(B) What are your estimations for these parameters?

(C) Suppose we have a new input vector X = (sunny, cold, high, strong, cool). What is $p\ (y = 1|X)$? Which class label will the Naive Bayes classifier assign to this example? Justify your answer.