**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.**

# Probability, MLE, and PAC [20 pts: 4+4+4+4+4]

* 1. Suppose that and are **independent** events, and . What is the value of ?
  2. Suppose that *X* and *Y* are **disjoint** events (i.e. ) and. What is the value of ?
  3. 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?*

* 1. 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 of having heads-up?

* 1. 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.

# Discriminant Linear Classifiers [20 pts: 10+10]

You are given a training data set of size *N* = 21. Each input vector is a point in the 2-dimensional Euclidean space *R*2. We have , , , , , , , , , , , , , , , , , , , , and . Each point is represented as a column vector.

There are two target classes C1 and C2. For each point in the training set, belongs to C1 if its second coordinate is less than or equal to 2, and belongs to C2 otherwise.

* 1. Compute the least-square linear classifier based on the training data (using 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 (a 3 by 2 matrix) [5pts].
  2. Compute the linear classifier based on the training data using Fisher’s linear discriminant by  where is the within-class covariance matrix. You need to write out (a) the Fisher criterion [5pts], (b) the computed parameter [5 pts].

# Continuous Bayes Classifier [20 pts: 5+5+5+5]

We want to build a Bayes classifier for a binary classification task ( or ) with a 1-dimensional input feature (). We know the following quantities: (1) ; (2) for and otherwise; and (3) for 0 ≤ *x* ≤ 8 and otherwise.

* 1. What is the prior for class label ?
  2. What is for ?
  3. For , what is the class label your classifier will assign? Why? What is the risk of this decision?
  4. What are the decision regions of your Bayes classifier?

# Discrete Bayes Classifier [20 pts: 5+5+5+5]

We want to build a Bayes classifier for a binary classification task ( or ) with input feature of two binary features ( and ). We know the following quantities: (1) ; (2) , , , , and (3) , ,.

* 1. What is the prior for class label ?
  2. What is
  3. For an example with and , what is the class label your classifier will assign? Why? What is the risk of this decision?
  4. What are the decision regions of your Bayes classifier?

# 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 , and (2) each column of being a binary feature.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Input Feature = (*, , , ,* ) | | | | | Class Label |
| **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

* 1. How many independent parameters are there in your Naive Bayes classifier? What are they (only list the independent parameters)? Justify your answer.
  2. What are your estimations for these parameters?
  3. Suppose we have a new input vector = (sunny*,* cold*,* high*,* strong*,* cool). What is ? Which class label will the Naive Bayes classifier assign to this example? Justify your answer.