

CSE 575: Homework #1

Due: September 27, 2022

Problem 1

- a) Suppose that X and Y are **independent** events, and $P(Y) > 0$. What is the value of $P(X|Y)$? (5pt)
- b) Suppose that X and Y are **disjoint** events, and $P(Y) > 0$. What is the value of $P(X|Y)$? (5pt)
- 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? (5pt)

HT, HT, TT, TT

- d) Suppose you are given a coin and are asked to toss as many times as you wish to decide the probability of getting heads from the coin toss. 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? (5pt)

Problem 2

You are given a training data set $\{x_n, t_n\}$ of size $N = 4$. Each input vector x_n is a point in the 2-dimensional Euclidean space R^2 . We have $x_1 = (1, 0)$, $x_2 = (2, 1)$, $x_3 = (2, 3)$, $x_4 = (3, 3)$.

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. If $x_n \in C_1$, we have $t_n = 1$. If $x_n \in C_2$, we have $t_n = 0$ in the equations regarding least-squares linear discriminant and Fisher's linear discriminant, and have $t_n = -1$ in the question on the perceptron algorithm.

Hint: you may use software like python, Matlab or R for matrix computation, especially inverses.

- a) Compute the least-square linear classifier based on the training data. You need to write out (a) the error function, (b) the computed parameters (w_0, w_1, w_2) . (6pt)
- b) Compute the linear classifier based on the training data using Fisher's linear discriminant. You need to write out (a) the error function, (b) the computed parameters (w_1, w_2) given $\|w\| = 1$, and (c) choose a threshold for $w^T x$ such that all the training examples are correctly classified. (6pt)
- c) Compute the linear classifier based on the training data using the perceptron algorithm, starting with the initial parameter $(w_0, w_1, w_2) = (1.5, 0, 0)$. For each iteration, you need to specify (a) the iteration number, (b) the current parameters, and (c) the updating vector. (8pt)

Problem 3

We want to build a Bayes classifier for a binary classification task ($y = 1$ or $y = 2$) with a 1-dimensional real-valued input feature (x). We know the following quantities: (1) $p(y = 1) = 0.4$; (2) $p(x|y = 1) = 0.5$ for $0 \leq x \leq 2$ and $p(x|y = 1) = 0$ otherwise; and (3) $p(x|y = 2) = 0.125$ for $0 \leq x \leq 8$ and $p(x|y = 2) = 0$ otherwise.

- a) What is the prior for class label $y = 2$? (5pt)
- b) What is $p(y = 1|x)$? (5pt)
- c) For $x = 1$, what is the class label your classifier will assign? Why? What is the risk of this decision (the probability of misclassifying)? (5pt)
- d) What is the decision function of your Bayes classifier? (describe the classification made for different values of x) (5pt)

Problem 4

We want to build a Bayes classifier for a binary classification task ($y = 1$ or $y = 2$) with two binary features (x_1 and x_2). We know the following quantities: (1) $P(y = 1) = 0.6$ (2) $P(x_1 = 0, x_2 = 0|y = 1) = 0.4$, $P(x_1 = 0, x_2 = 1|y = 1) = 0.3$, $P(x_1 = 1, x_2 = 0|y = 1) = 0.2$, $P(x_1 = 1, x_2 = 1|y = 1) = 0.1$, and (3) $P(x_1 = 0, x_2 = 0|y = 2) = 0.3$, $P(x_1 = 0, x_2 = 1|y = 2) = 0.1$, $P(x_1 = 1, x_2 = 0|y = 2) = 0.4$, $P(x_1 = 1, x_2 = 1|y = 2) = 0.2$

- a) What is the prior for class label $y = 2$? (5pt)
- b) What is $P(y = 1|x)$? (5pt)
- 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? (5pt)
- d) What is the decision function of your Bayes classifier? (5pt)

Problem 5

Given the training data below, we want to train a binary classifier using Naive Bayes. The last column is the class label y , and each column of x denotes a categorical feature. Features **Sky**, **Humid**, and **Wind** have two possible values; feature **Temp** can take three values (cold, mild, hot).

input features $x = (x_1, x_2, x_3, x_4)$				Class label y
Sky	Temp	Humid	Wind	Play Sport
sunny	mild	normal	strong	yes
rainy	cold	high	mild	no
rainy	hot	normal	strong	no
sunny	hot	high	mild	no
sunny	cold	normal	mild	yes
sunny	mild	normal	strong	no
rainy	mild	high	strong	no
rainy	mild	normal	mild	yes
sunny	hot	normal	strong	yes
sunny	cold	high	strong	no

- a) How many independent parameters are there in your Naive Bayes classifier? Justify your answer. (5pt)
- b) What are the maximum likelihood estimates for these parameters? (10pt)
- c) Suppose we have a new input vector $\mathbf{x} = (\text{sunny}, \text{cold}, \text{normal}, \text{strong})$. What is $P(y = 1|\mathbf{x})$? Which class label will the Naive Bayes classifier (with the parameters from part (b)) assign to this example? Justify your answer. (5pt)
- d) (Bonus question) Suppose we have a new input vector $\mathbf{x} = (\text{sunny}, \text{cold}, \text{normal}, \text{missing})$. That is, you are given values for the first three features only. What is $P(y = 1|\mathbf{x})$ according to the Naive Bayes model from part (b)? Justify your answer. (5pt)