Q2 - ML Basics & Baye's Classifier

Total points 20/20

1. Upload /Submit solutions for (b) parts of Q 1, 3 and 4.
2. Note that the Questions are shuffled.

Max Marks: 20
Max Time : 35 mts (+15 minutes if you are uploading three files)

✓ Q3. (a) What is the number of parameters needed to represent a Naive Bayes classifier with n Boolean features and a Boolean label ? 1/1

- (●) 2n + 1 ✓
- ( ) n + 1
- ( ) 2n
- ( ) n
- ( ) Other: .....

✓ Q4. a. Mark the correct option. 1/1

Consider the following three training sets with two features and two classes.

T1

X1	X2	C
0	0	0
1	1	1
1	0	1
0	1	1

T2

X1	X2	C
1	0	0
0	0	0
1	1	1
0	1	0

T3

X1	X2	C
1	1	0
0	1	1
1	0	1
0	0	0

Assume that the hypothesis space consists of linear decision boundaries. The number of separating lines that can be drawn for each of the training sets is

a) exactly one, exactly one, none, respectively
b) infinite, infinite, infinite respectively
c) exactly one, exactly one, infinite, respectively
d) infinite, infinite, none, respectively

- ( ) A
- ( ) B
- ( ) C
- (●) D ✓

✓ Q 3(b). List all the parameters of the NB classifier in Q 3(a), Assume class labels as C1 and C2, and features as F1, ..., Fn. 3/3

Prior P(C1) (1 probability, as P(C2) = 1-P(C1))), and class conditionals P(F1|C1), P(F2|C1) ... P(Fn|C1) (n probabilities) and P(F1|C2), P(F2|C2), ... P(Fn|C2) (n probabilities)

✓ Q2. Let Boolean random variables LC denote "has lung cancer" and T stand for " test is positive." From historical data it is known that. (a) The prior probability of having lung cancer is 0.01. (b) The probability of testing positive for lung cancer is 90%. (c) The probability of testing negative when the patient does not have lung cancer is 89.9%. FIND (A) probability of having a positive test, (B) If a patient has a positive test, what is the probability that he/she has lung cancer? 6/6

✓ Q1. b. Show the area corresponding to Baye's error for two-class problem with one attribute. Assume that the classes follow Gaussian [~N(Mu(i), Sigma-sq(i))]. Show the point "x" mentioned in Q1. a. Label the diagram neatly, clearly showing notation. 5/5

✓ Q1. a. In a two class classfication problem, a point x on the Bayes optimal decision boundary always satisfies 1/1

- (●) P(C1| x) = P(C2 | x) ✓
- ( ) P(C1| x) != P(C2 | x)
- ( ) P(C1| x) >= P(C2 | x)
- ( ) P(C1| x) <= P(C2 | x)

✓ Q4. b. Draw the decision boundaries for all three training sets in Q 4. a. 3/3