Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division
Second Semester 2019-20
M.Tech. (Data Science and Engineering)
Comprehensive Examination (Regular)

Course No. : DSECLZG565
Course Title : MACHINE LEARNING
Nature of Exam : Open Book
Weightage : 40%
Duration : 2 Hours
Date of Exam: July 12, 2020 Time of Exam: 10:00 AM – 12:00 PM

No. of Pages = 3
No. of Questions = 5

Note: Assumptions made if any, should be stated clearly at the beginning of your answer.

Question 1. **[3+3+2+3=11 marks]**

A) Suppose you flip a coin with unknown bias θ; P(x = H | θ) = θ, five times and observe the outcome as HHHHH.
What is the maximum likelihood estimator for θ? **[1 mark]**
Would you think this is a good estimator? If not, why not? **[2 marks]**

Solution :
Section 6.5 of Tom Mitchell's book for 1(a).
Section 6.9.1.1 of Tom Mitchell's book for 1(b) and 1(c)
What is the maximum likelihood estimator for ϑ?
Using normal approach
^
θ = 5 / 5 = 5 / 5 = 1
*Would you think this is a good estimator? If not, why not? **[2 marks]***
Section 6.9.1.1 of Tom Mitchell's book
Not a good estimate.
Because of number of observations for x = T (observation of T) being zero, above mechanism produces biased overestimate of probability for x = H (and biased underestimate of probability of x = T). Due to this, any future query based on this would result into either overestimate or underestimate.

B) A disease has four symptoms and past history of a physician has the following data. Use Naïve Bayes classifier to predict whether patient has disease for new patient data symptoms. **[2 marks]**

	Symp1	Symp2	Symp3	Symp4	Disease
1	yes	no	mild	yes	no
2	yes	yes	no	no	yes
3	yes	no	strong	yes	yes
4	no	yes	mild	yes	yes
5	no	no	no	no	no
6	no	yes	strong	yes	yes
7	no	yes	strong	no	no
8	yes	yes	mild	yes	yes

For a new patient				
Symp1	Symp2	Symp3	Symp4	Disease
yes	no	mild	yes	?

Solution:	1	
Prior	0.625	0.375
	Dis	No dis
	P(s/D)	P(s/ND)
s1=yes	0.6	0.333
s2=no	0.2	0.666
s3=mild	0.4	0.333
s4=yes	0.8	0.333
	0.0384	0.024593
Posterior	0.024	0.009222
Disease=Yes		

C)

1. Can logistic regression be applied to multi-class classification problem? State true or false **[1 mark] true**
2. Why are log probabilities computed instead of probabilities? **[1 mark]**
   a. To make computation consistent
   b. To factor into smaller values of probabilities
   c. To factor into larger values of probabilities-correct
   d. None of these

D)

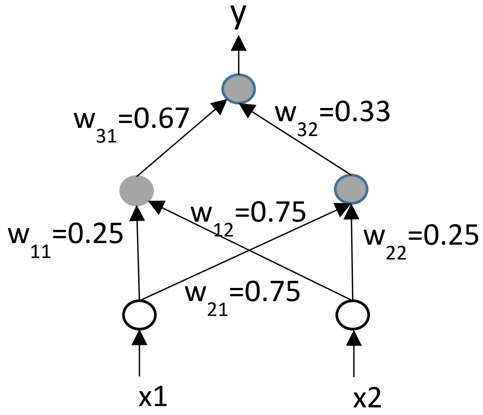
1. In a linear relationship y = m*x+b, y is said to be dependent on x when: **[1 mark]**
   a. m is closer to zero.
   b. m is far from zero.
   c. b is far from zero.
   d. b is closer to zero.

2. In a linear relationship between y and x, y is not dependent on x when: **[1 mark]**
   a. The coefficient is closer to zero.
   b. The coefficient is far from zero.
   c. The intercept is far from zero.
   d. The intercept is closer to zero.

3. In a linear regression model y= w0 + w1*x, if true relationship between y and x is y = 7.5 +3.2x, then w0 acts as, **[1 mark]**
   a. Intercepts
   b. Coefficients
   c. Estimators
   d. Residuals

Question 2.

The following backpropagation network uses an activation function called leaky ReLU that generates output = input, if input >= 0, and 0.1 * input if output < 0. At a particular iteration, the weights are indicated in the following figure. Training error is given by E = $0.5*(t-y)^2$ where t is the target output and y is the actual output from the network. What are the outputs of hidden nodes and actual final output y from the network with x1=x2=1? What will be the weights w_{31} and w_{12} in the next iteration with learning rate = 0.1, x1=x2=1, and target output t=0? Assume derivative of activation function = 0 at input = 0, and zero bias at all nodes. [1+1+1+1.5+2.5=**7 marks]**



For x1=x2=1, actual output y = 1, and target output t = 0. Output of hidden nodes are o_1=o_2=1. So, error $E = 0.5*(t\text{-}y)^2$

So,

$$\Delta w_{31} = -\frac{\eta \delta E}{\delta w_{31}} = -\frac{\eta \delta E}{\delta y} * \frac{\delta y}{\delta w_{31}} = \eta(t - y)y * o_1 = -0.1.$$

Thus, value of w_{31} in the next iteration will be

$$\mathrm{w}_{31} + \Delta w_{31} = 0.67 - 0.1 = 0.57.$$

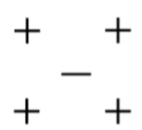
$$\Delta w_{12} = -\frac{\eta \delta E}{\delta w_{12}} = -\frac{\eta \delta E}{\delta y} * \frac{\delta y}{\delta w_{12}} = -0.1 \; \frac{\delta y}{\delta o_1}\frac{\delta o_1}{\delta w_{12}} = -0.1 * w_{31} * x_2 = -0.067\ .$$

Thus, value of w_{12} in the next iteration will be

$$\mathrm{w}_{12} + \Delta w_{12} = 0.75 - 0.067 = 0.683.$$

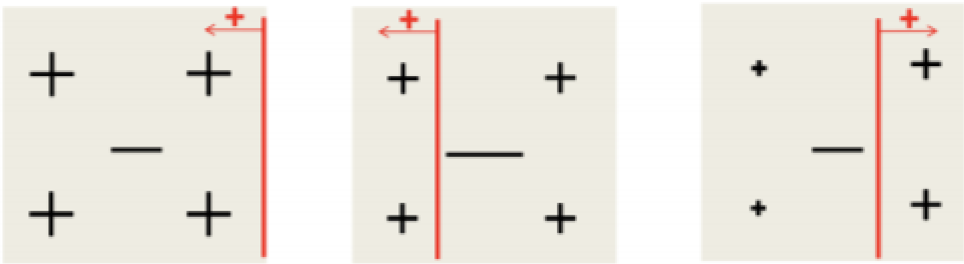
Question 3.

A) Consider training a boosting classifier using decision stumps on the following data set:



1. Circle the examples which will have their weights increased at the end of the first iteration? **[2 marks]**
2. How many iterations will it take to achieve zero training error? Explain**. [3 marks]**
Solution

1. The negative example since the decision stump with least error in first iteration is constant over the whole domain. Notice this decision stump only predicts incorrectly on the negative example, whereas any other decision stump predicts incorrectly on at least two training examples.



2. At least three iterations. The first iteration misclassifies the negative example, the second iteration misclassifies two of the positive examples as the negative one has large weight. The third iteration is needed since a weighted sum of the first two decision stumps can't yield zero training error, and misclassifies the other two positive examples. See Figures below (Fig. 5)

B) A new mobile phone service chain store would like to open 20 service centres in Bangalore. Each service centre should cover at least one shopping centre and 5,000 households of annual income over 75,000. Design a scalable algorithm that decides locations of service centres by taking all the aforementioned constraints into consideration **[5 marks]**

The algorithm would be:
1. Identify the households (their location) with annual income over 75,000.
2. Identify shopping centers (their location) in the city.
3. Use k-Means to form 20 (k=20) clusters of shopping centers based on their distances with the other shopping centers.
4. Take households as query points to insert into these clusters formed above.

4a. An household is assigned to the cluster by comparing two closest shopping centers, and one with the less number of households attached to it is chosen. (We decided to attach households to either of the two nearest shopping centers as a user is less prone to go to the third distant service center leaving the first two.)

Question 4.

In a clinical trial, height and weight of patients is recorded as shown below in the table. For incoming patient with weight = 58 Kg and Height = 180 cm, classify if patient is Under-weight or Normal using KNN algorithm with When K = 3? [**5 marks]**

Weight (in Kg)	Height (in cm)	Class
61	190	Under-weight
62	182	Normal
57	185	Under-weight
51	167	Under-weight
69	176	Normal
56	174	Under-weight
60	173	Normal
55	172	Normal
65	172	Normal

Question 5.

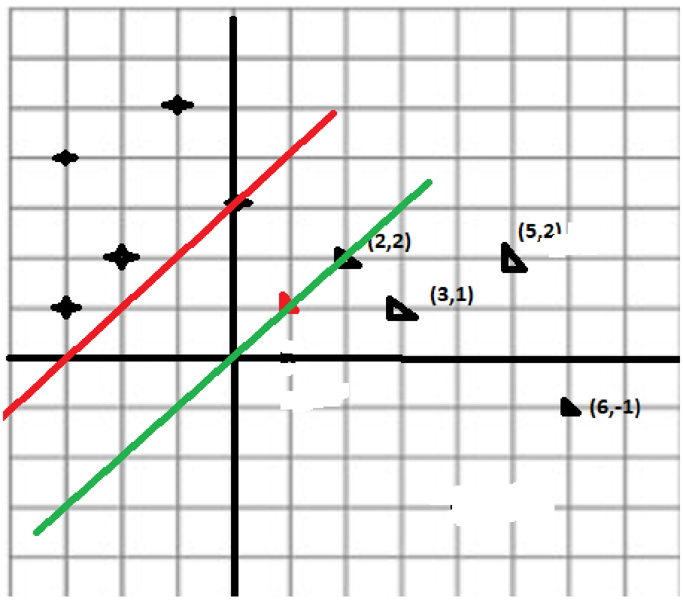
Considering the following data, Let x1, x2 be the features

Positive Points: {(3, 1), (5, 2), (1, 1), (2, 2), (6, -1)}

Negative Points: {(-3, 1), (-2, 2), (0, 3), (-3, 4), (-1, 5)}

Derive an equation of hyperplane and compute the model parameters**. [7 marks]**

Let x1, x2 be the features



The positive examples (triangles): (1, 1) and (2, 2) as these are extreme points with respect to rest of the examples on positive side.

Identification of support verctors -> 1M

The equation of hyperplane corresponding to support vectors of + examples:

x1 –x2 =0

Similarly, (3,0) is one support vector on the –ve side of the decision boundary.
and Let equation of decision boundary will be of the form: cx1 –cx2 +A=0

The equation of the hyperplane through –ve example: x1-x2+A=0,
Since it passes through (0, 3) => A= 3

Equation of hyperplane=x1 –x2 +3=0 → 3M
Margin is given by distance between parallel lines x1-x2=0 and x1-x2+3=0 is given by 3/√*2* ***3M***

Model parameters:

The equation of decision boundary can be Cx1 -Cx2 +B=0.

Width of the decision boundary d =2/|w|
2/|w| =3/√*2*

- ⇨ 2/√*2C =3/√2*
- ⇨ C=2/3
- ⇨ **Model parameters are [2/3, 2/3] Model parameters: 3M**