Birla Institute of Technology & Science, Pilani
Work-Integrated Learning Programmes Division
MTech. Software Engineering at DSE (FC04, FA04_1-2021) Cluster

Second Semester 2021-2022
Mid-Semester Test
(EC-2 Regular)

Course No.	: DSECLZG565
Course Title	: Machine Learning
Nature of Exam	: Open Book
Weightage	: 30%
Duration	: 2 Hours
Date of Exam	: 10-07-2022(FN)

No. of Pages = 2
No. of Questions = 6

Note:
1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Q.1 Let $T_1, T_2, \ldots T_n$ *be* a random sample of a population describing the website loading time on a mobile browser with probability density function given as:

$$f\left(t/\theta\right) = \frac{1}{\theta} t^{\frac{(1-\theta)}{\theta}} \quad where \;\; 0 < t < 1 \; and \; 0 < \theta < \infty$$

Find the maximum likelihood estimator of θ. What is the estimate of θ, if the website loading time from four samples are $t_1 = 0.10$, $t_2 = 0.22$, $t_3 = 0.54$, $t_4 = 0.36$. [5 Marks]

Solution:

Q1 Solution :-

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} x_i^{(1-\theta)/\theta}$$

$$= \theta^{-n} \left(\prod_{i=1}^{n} x_i \right)^{\left(\frac{1-\theta}{\theta}\right)}$$

$$\log L(\theta) = -n \log \theta + \frac{1-\theta}{\theta} \sum_{i=1}^{n} \log x_i$$

$$= -n \log \theta + \frac{1}{\theta} \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \log x_i$$

$$= \frac{d}{d\theta} \log L(\theta) = \frac{-n}{\theta} - \frac{1}{\theta^2} \sum_{i=1}^{n} \log x_i = 0$$

$$\hat{\theta} = -\frac{1}{n} \sum_{i=1}^{n} \log x_i$$

Now we have the estimator, and for given data, the estimate value is

$$\hat{\theta} = \frac{-1}{n} \sum_{i=1}^{n} \log x_i$$

$$= \frac{-1}{4} \log (0.10 \cdot 0.22 \cdot 0.54 \cdot 0.36)$$

$$\boxed{1.3636}$$

Marking Scheme: Derivation of θ =3 marks (step wise marks)

θ Computation = 2 marls (wrong value = 0 marks)

Q.2 As a part of efforts to improve students' performance in the exams, you have been given the data showing number of study hours spent by students, their gender and their final results as pass or fail. Using this sample dataset, apply Naïve Bayes classification technique, to classify the test case {No of study hours = 3.5, Gender="male"} either as "Pass", or "Fail". [5 Marks]

No of study hours	Gender	Final result
4.5	Male	Pass
7	Female	Pass
2	Male	Fail
4	Female	Fail
2.5	Male	Fail
3	Female	Fail
8.3	Male	Fail
8	Female	Pass
9	Male	Pass

Solution:

1. Prior: [1M]

p(y=Pass)	p(y=Fail)
0.444444	0.555556

2. No of study hours –X1: continuous variable, applying class conditional PDF [1M]
3.

$$\boxed{p(X_i = x|Y = y_k) = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} e^{-\frac{1}{2}(\frac{x-\mu_{ik}}{\sigma_{ik}})^2}}$$

	Variance	mean
Pass class	2.945	7.2
Fail class	4.64	3.9

4. X1=3.5, X2="male" [3M]

$$\hat{Y} \leftarrow \underset{y_k}{\operatorname{argmax}} P(Y = y_k)\Pi_i P(X_i|Y = y_k)$$

p(X1/ y=Pass)	0.105614
p(X1/ y=Fail)	0.184564

p(X2/ y=Pass)	0.5
p(X2/ y=Fail)	0.6

P(y=Pass/X)	0.02346969
P(y=Fail/X)	0.061521395

Class : Fail

Q.3 The 2-input AND gate is implemented using logistic regression classifier with gradient descent optimization algorithm. The model parameters at time *t* are given by θ_0=0, θ_1=0, and θ_2=0. Given binary input (x1,x2), [2+3 = 5 Marks]

a) What will be value of the loss function at t? [2M]

Solution:

Cross entropy loss:

$$= -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))]$$

x1	x2	Target y	Actual Output-yhat	y.ln(yhat)+(1-y)ln(1-yhat)
0	0	0	0.5	0*ln0.5+(1-0)*ln(1-0.5)
0	1	0	0.5	0*ln0.5+(1-0)*ln(1-0.5)
1	0	0	0.5	0*ln0.5+(1-0)*ln(1-0.5)
1	1	1	0.5	1*ln(0.5)

total loss=	0.693147181

b) What will be the values of θ_0, θ_1 and θ_2 at (t+1) with learning rate α=1 and L2 regularization constant λ=1? [3M]

Solution:
Cost function

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

Apply gradient descent update rule

y-hat	y	yhat-y	x0	(yhat-y)x0	w0-new
0.5	0	0.5	1	0.5	-0.25
0.5	0	0.5	1	0.5	
0.5	0	0.5	1	0.5	
0.5	1	-0.5	1	-0.5	

y-hat	y	yhat-y	x1	(yhat-y)x1	regularized w1-new
0.5	0	0.5	0	0	0
0.5	0	0.5	0	0	
0.5	0	0.5	1	0.5	
0.5	1	-0.5	1	-0.5	

y-hat	y	yhat-y	x2	(yhat-y)x1	regularized w2-new
0.5	0	0.5	0	0	0
0.5	0	0.5	1	0.5	
0.5	0	0.5	0	0	
0.5	1	-0.5	1	-0.5	

Q.4 We claim that there exists a value for α in the following data : (1.0, 4.0), (2,0, 9.0) , (3.0, α) such that the line $y = 2 + 3x$ is the best least-square fit for the data. Is this claim true? If the claim is true, find the value of α. Otherwise, explain why the claim is false. Give detailed mathematical justification for your answer. [5 Marks]

For the line $y = a + bx$, the MSE turns out to be

$$(4.0 - (a+b))^2 + (9.0 - (a+2b))^2 + (\alpha - (a+3b))^2$$

$$\frac{\partial MSE}{\partial a} = 0$$

$$\Rightarrow 2(4.0 - a - b)(-1) + 2(9.0 - a - 2b)(-1) + 2(\alpha - a - 3b)(-1) = 0$$

$$\Rightarrow 4.0 - a - b + 9.0 - a - 2b + \alpha - a - 3b = 0$$

$$\alpha + 13 - 3a - 6b = 0$$

$$\Rightarrow \alpha = 3a + 6b - 13 \quad - \ 1$$

$$\frac{\partial MSE}{\partial b} = 0$$

$$\Rightarrow 2(4.0 - a - b)(-1) + 2(9.0 - a - 2b)(-2) + 2(\alpha - a - 3b)(-3) = 0$$

$$4.0 - a - b + 18 - 2a - 4b + 3\alpha - 3a - 9b = 0$$

$$22 + 3\alpha - 6a - 14b = 0$$

$$\frac{6a + 14b - 22}{3} = \alpha \quad - \ 2$$

From ① and ②, the parameters a and b for the best fit must solve

$$\frac{6a + 14b - 22}{3} = 3a + 6b - 13$$

$$\Rightarrow 6a + 14b - 22 = 9a + 18b - 39$$

$$3a + 4b = 17$$

Substituting $a = 2$ and $b = 3$, we see that the above requirement is not met $\Rightarrow$ the given claim is false

Marking Scheme: calculation of 1 and 2 – 3M
Equation of a and b = 1M
Final answer =1M

Q.5 Consider a basis function $\phi_j(x) = x^j$, which is used to model nonlinear function of the input variables of the form $y(x,\theta) = \sum_{j=0}^{2} \theta_j \phi_j(x)$. Determine θ_0, θ_1 and θ_2 for the table given below. [6 Marks]

x	y
0	1
1	3
2	7
5	31

Solution:

Polynomial Regression: $y = \theta_0 + \theta_1 x + \theta_2 x^2$ [2M]

Solution: Method 1 **[4M]**

Q.5 Consider a basis function $\varphi_j(x) = x^j$, which is used to model nonlinear function of the input variables of the form $y(x,\theta) = \sum_{j=0}^{2} \theta_j \varphi_j(x)$. Determine θ_0, θ_1 and θ_2 for the table given below.

x	y
0	1
1	3
2	7
5	31

Solution 1: Using Polynomial Regression:

$$y(x,\theta) = \sum_{j=0}^{2} \theta_j \varphi_j(x) \qquad [\varphi_j(x) = x^j]$$

$$y \Rightarrow \theta_0 x^0 + \theta_1 x^1 + \theta_2 x^2$$

$$\boxed{y \Rightarrow \theta_0 + \theta_1 x + \theta_2 x^2}$$

Use the following equation to determine θ_0, θ_1 and θ_2

$$n a_0 + (\Sigma x_i) a_1 + (\Sigma x_i^2) a_2 = \Sigma y_i \rightarrow (1)$$

$$(\Sigma x_i) a_0 + (\Sigma x_i^2) a_1 + (\Sigma x_i^3) a_2 = \Sigma x_i y_i \rightarrow (2)$$

$$(\Sigma x_i^2) a_0 + (\Sigma x_i^3) a_1 + (\Sigma x_i^4) a_2 = \Sigma x_i^2 y_i \rightarrow (3)$$

	x	y	x^2	x^3	x^4	$x*y$	x^2*y
$n=4$	0	1	0	0	0	0	0
	1	3	1	1	1	3	3
	2	7	4	8	16	14	28
	5	31	25	125	625	155	775
Sum →	8	42	30	134	642	172	806

Here

$\Sigma x_i = 8$

$\Sigma y_i = 42$

$\Sigma x_i^2 = 30$

$\Sigma x_i^3 = 134$

$\Sigma x_i^4 = 642$

$\Sigma x_i y_i = 172$

$\Sigma x_i^2 * y_i = 806$

$n = 4$

Substitute the above values in all 3 equation

$$4a_0 + 8a_1 + 30a_2 = 42$$
$$8a_0 + 30a_1 + 134a_2 = 172$$
$$30a_0 + 134a_1 + 642a_2 = 806.$$

Coeffient Matrix. Constant Matrix

$$\begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array}\begin{bmatrix} 4 & 8 & 30 \\ 8 & 30 & 134 \\ 30 & 134 & 642 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 42 \\ 172 \\ 806 \end{bmatrix}$$

Step 1 :- $R_1 \rightarrow R_1/4$

$$\begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array}\left[\begin{array}{ccc|c} 1 & 2 & 7.5 & 10.5 \\ 8 & 30 & 134 & 172 \\ 30 & 134 & 642 & 806 \end{array}\right]$$

[R_{11} ($4/4 \Rightarrow 1$) R_{12} ($8/4 \Rightarrow 2$)
R_{13} ($30/4 \Rightarrow 7.5$)
$C_1 \Rightarrow 42/4 \Rightarrow 10.5$]

Step 2 :- $R_2 \rightarrow R_2 - R_{21} * R_1$

$$\begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array}\left[\begin{array}{ccc|c} 1 & 2 & 7.5 & 10.5 \\ 0 & 14 & 74 & 88 \\ 30 & 134 & 642 & 806 \end{array}\right]$$

[$R_{21} \rightarrow (8 - 8*1 \Rightarrow 8-8 \Rightarrow 0)$
$R_{22} \rightarrow (30 - 8*2 \Rightarrow 30-16 \Rightarrow 14)$
$R_{23} \Rightarrow (134 - 8*7.5 \Rightarrow 134 - 60$
$\Rightarrow 74$
$C_2 \Rightarrow (172 - 8*10.5 \Rightarrow 172 - 84$
$\Rightarrow 88$]

$$\begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array}\left[\begin{array}{ccc|c} 1 & 2 & 7.5 & 10.5 \\ 0 & 14 & 74 & 88 \\ 0 & 74 & 417 & 491 \end{array}\right]$$

step 4 :- $R_2 \rightarrow R_2 / R_{22}$

$$\begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array}\left[\begin{array}{ccc|c} 1 & 2 & 7.5 & 10.5 \\ 0 & 1 & 5.285714 & 6.285714 \\ 0 & 74 & 417 & 491 \end{array}\right]$$

$R_{21} \Rightarrow 0/14 \Rightarrow 0$

$R_{22} \Rightarrow 14/14 \Rightarrow 1$

$R_{23} \Rightarrow 74/14 \Rightarrow 5.285714$

$C_2 \Rightarrow 88/14 \Rightarrow 6.285714$

step5: $R_3 \rightarrow R_3 - R_{32} * R_2$

$$\begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array}\left[\begin{array}{ccc|c} 1 & 2 & 7.5 & 10.5 \\ 0 & 1 & 5.285714 & 6.285714 \\ 0 & 0 & 25.85714 & 25.85714 \end{array}\right]$$

$R_{31} \Rightarrow 0 - 74 * 0 \Rightarrow 0$

$R_{32} \Rightarrow 74 - 74 * 1 \Rightarrow 0$

$R_{33} \Rightarrow 417 - 74 * 5.285714 \Rightarrow 25.85714$

$C_3 \Rightarrow 491 - 74 * 6.285714 \Rightarrow 25.85714$.

$\Downarrow$

$$1 * a_0 + 2 * a_1 + 7.5 * a_2 = 10.5 \rightarrow (1)$$

$$1 * a_1 + 5.285714 * a_2 = 6.285714 \rightarrow (2)$$

$$25.85714 * a_2 = 25.85714 \rightarrow (3)$$

Solve the above equation (3)

$25.85714(a_2) = 25.85714$

$\boxed{a_2 = 1}$

substitute $a_2 = 1$ in equation (2)

$a_1 + 5.285714 * 1 = 6.285714$

$a_1 = 6.285714 - 5.285714$

$\boxed{a_1 = 1}$

substitute $a_1 = 1$ & $a_2 = 1$ in equation (1)

$a_0 + 2 * 1 + 7.5 * 1 = 10.5$

$a_0 + 9.5 = 10.5$

$a_0 = 10.5 - 9.5$

$\boxed{a_0 = 1}$

$\boxed{a_0 = 1 \; ; \; a_1 = 1 \; ; \; a_2 = 1}$

Solution: Method 2

Using closed form solution: [4M]

$$\boldsymbol{\theta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

$\theta_0 = 1$, $\theta_1 = 1$ and $\theta_2 = 1$

Q.6 Consider the dataset of binary values in terms of attribute-value pairs where F is the value, and A,B, C are attributes. What is the entropy of the dataset? Fill in the columns for A and B, if it is known that A has maximum information gain and B has minimum information gain. Give mathematical justification for your answer. [4 Marks]

A	B	C	F
		0	0
		1	1
		0	1
		1	0
		0	1
		1	1
		0	1
		1	1

Solution:
For the entropy problem, the column F is the output attribute.

2 Marks:
Let column A = column F, so that the ith entries of the two column match each other. The information gain can be written as
$InformationGain(S, A) = Entropy(S) - \sum \frac{|S_A|}{|S|} Entropy(S_A)$ where the sum is over the attribute values of A.
Since the column entries of A match with those of F we see that the set $S_{A=0}$ is full of 0s and $S_{A=1}$ is full of 1s, so that $Entropy(S_{A=0})$ is 0 and so is $Entropy(S_{A=1})$. From the equation on Information Gain we can see that we get the maximum information gain possible in this case.

2 Marks:
The information gain with respect to column B can be written as $InformationGain(S, B) = Entropy(S) - \sum \frac{|S_B|}{|S|} Entropy(S_B)$.
For minimum information gain we see that if let the column B be the column of all 1s, then we have $S_{B=1} = S$ and $S_{B=0} = \phi$. Once again plugging this into the information gain equation shows that the information gain with respect to B is 0.

The arguments above work for maximum information gain when A is taken to be complement of F, and B is taken to be all zeroes rather than all 1s.