**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

No. of Pages = 2

No. of Questions = 6

Nature of Exam : Open Book

Weightage : 30%

Duration : 2 Hours

Date of Exam : 10-07-2022(FN)

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.
4. Let *T1,T2, …. Tn be* a random sample of a population describing the

website loading time on a mobile browser with probability density

function given as:

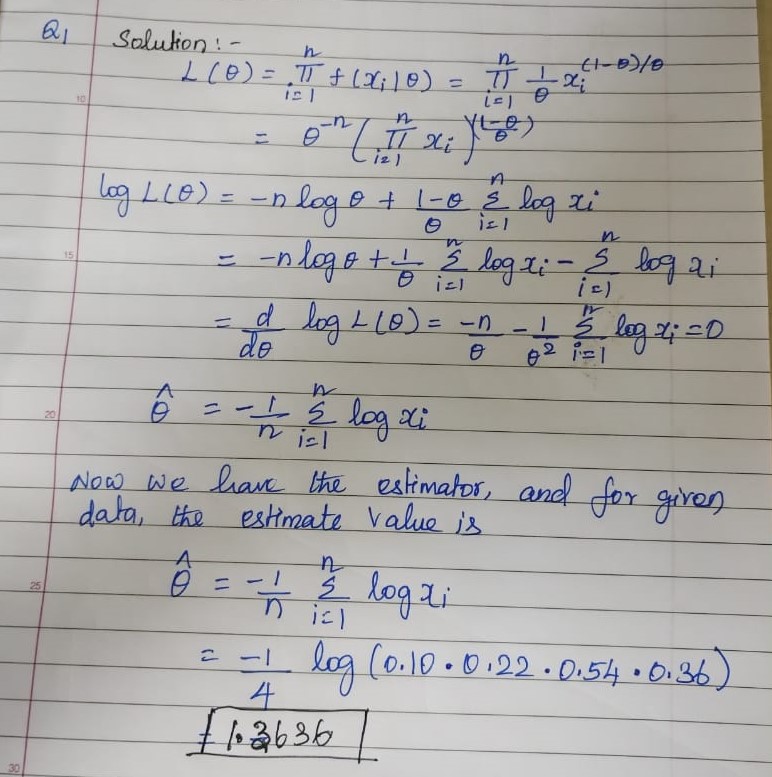


Find the maximum likelihood estimator of*θ*. What is the estimate of *θ,*

if the website loading time from four samples are t1 = 0.10, t2 = 0.22, t3

= 0.54, t4 = 0.36. [5 Marks]

Solution:



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

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

1. 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 |

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



|  |  |  |
| --- | --- | --- |
|  | Variance | mean |
| Pass class | 2.945 | 7.2 |
| Fail class | 4.64 | 3.9 |
|  |  |  |

1. X1=3.5, X2=”male” [3M]

|  |  |
| --- | --- |
| 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

1. The 2-input AND gate is implemented using logistic regression classifier with

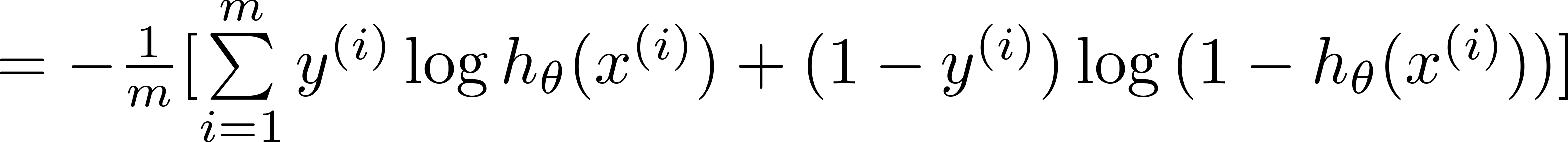
gradient descent optimization algorithm. The model parameters at time *t* are given

by θ0=0, θ1=0, and θ2=0. Given binary input (x1,x2), [2+3 = 5 Marks]

1. What will be value of the loss function at t? [2M]

Solution:

Cross entropy loss:



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

1. What will be the values of θ0, θ1 and θ2 at (t+1) with learning rate α=1 and L2 regularization constant λ=1? [3M]

Solution:

Cost function



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 |  |

1. We claim that there exists a value for in the following data : (1.0, 4.0), (2,0, 9.0) ,

(3.0, such that the line is the best least-square fit for the data. Is

this claim true? If the claim is true, find the value of. Otherwise, explain why the

claim is false. Give detailed mathematical justification for your answer. [5 Marks]



Marking Scheme: calculation of 1 and 2 – 3M

Equation of a and b = 1M

Final answer =1M

1. Consider a basis function, which is used to model nonlinear function of

the input variables of the form  . Determine θ0, θ1 and θ2 for

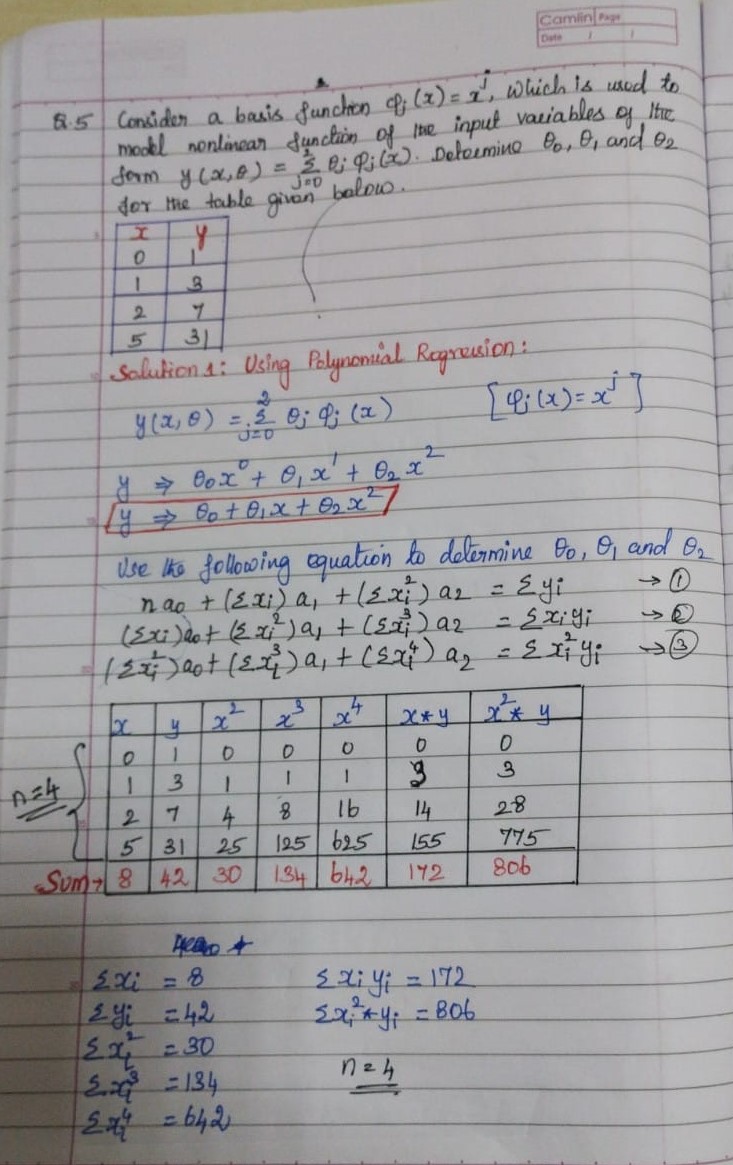
the table given below. [6 Marks]

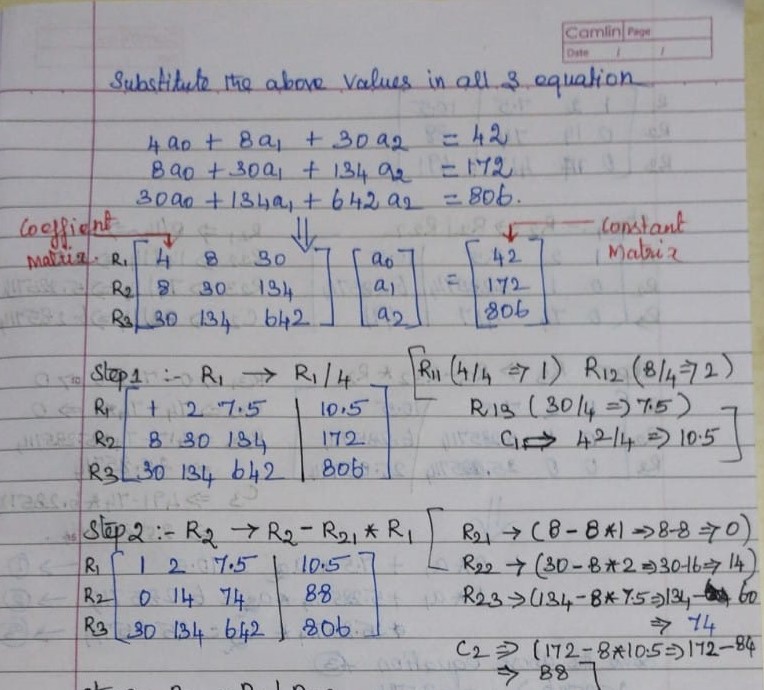
|  |  |
| --- | --- |
| x | y |
| 0 | 1 |
| 1 | 3 |
| 2 | 7 |
| 5 | 31 |

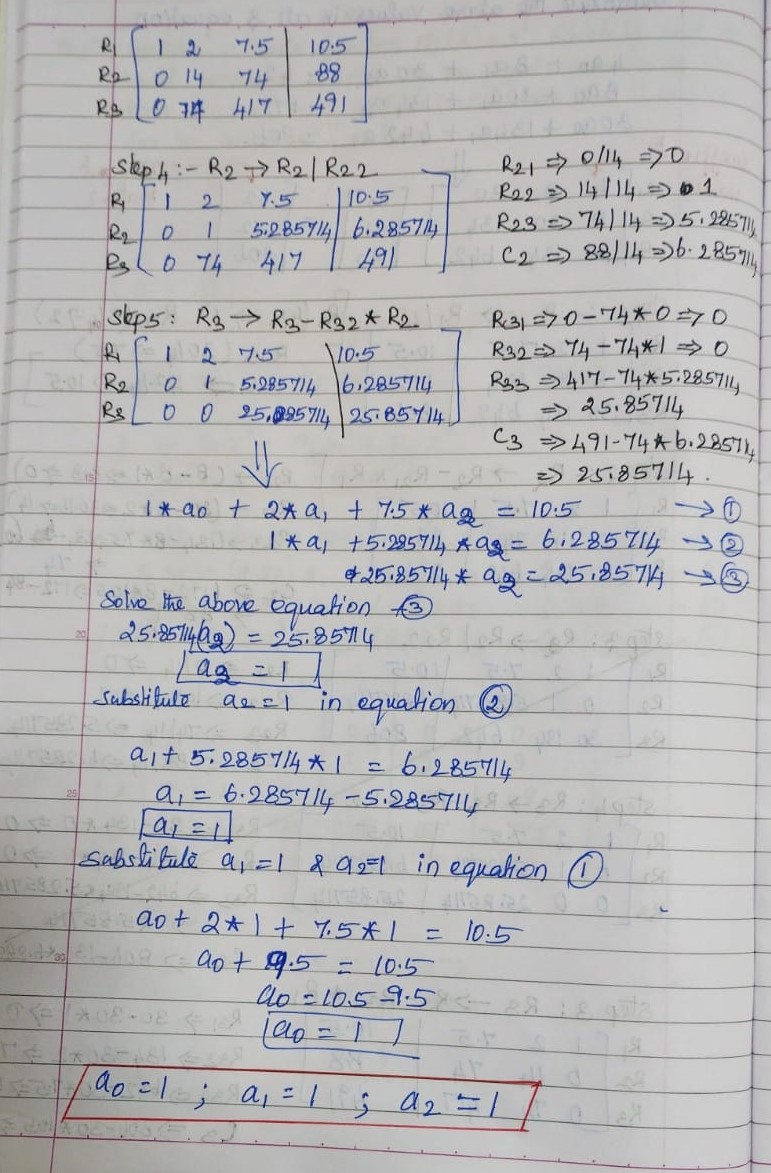
Solution:

Polynomial Regression:  [2M]

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

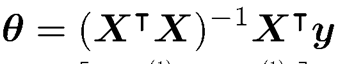






**Solution: Method 2**

Using closed form solution: [4M]



θ0 = 1, θ1 = 1 and θ2 =1

1. 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 th entries of the two column match each other. The information gain can be written as

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 is full of 0s and is full of 1s, so that is 0 and so is . 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

For minimum information gain we see that if let the column B be the column of all 1s, then we have and . 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.**