**(EC-2 Makeup)**

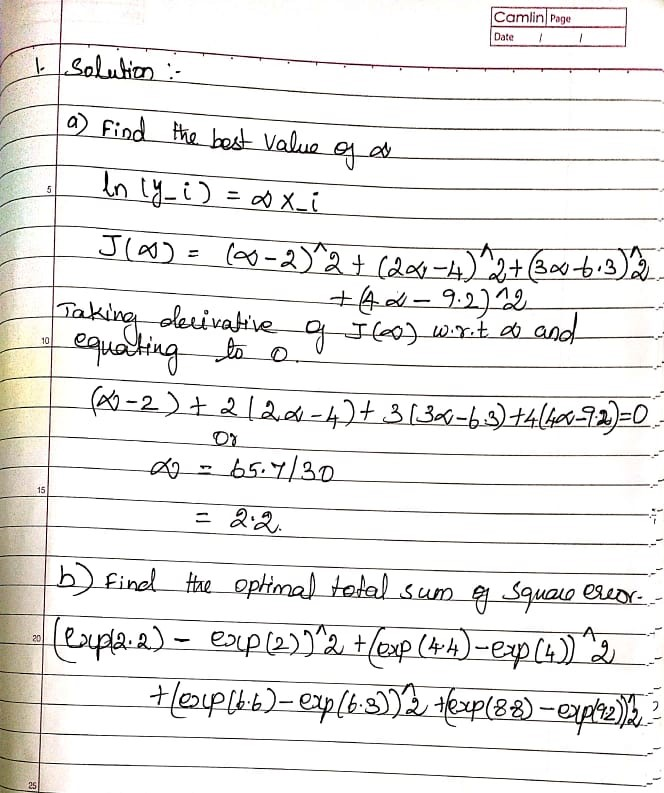
1. Consider the following dataset with 4 records. [4+2 = 6 Marks]

|  |  |
| --- | --- |
| Input X | Output Y |
| 1 | exp(2) |
| 2 | exp(4) |
| 3 | exp(6.3) |
| 4 | exp(9.2) |

Assume output y=e(α \* x) . Using linear regression,

1. Find the best value of α.
2. Find the optimal total sum of square error.

Solution:



1. Consider inputs *xi* which are real valued attributes and the outputs *yi* which are real

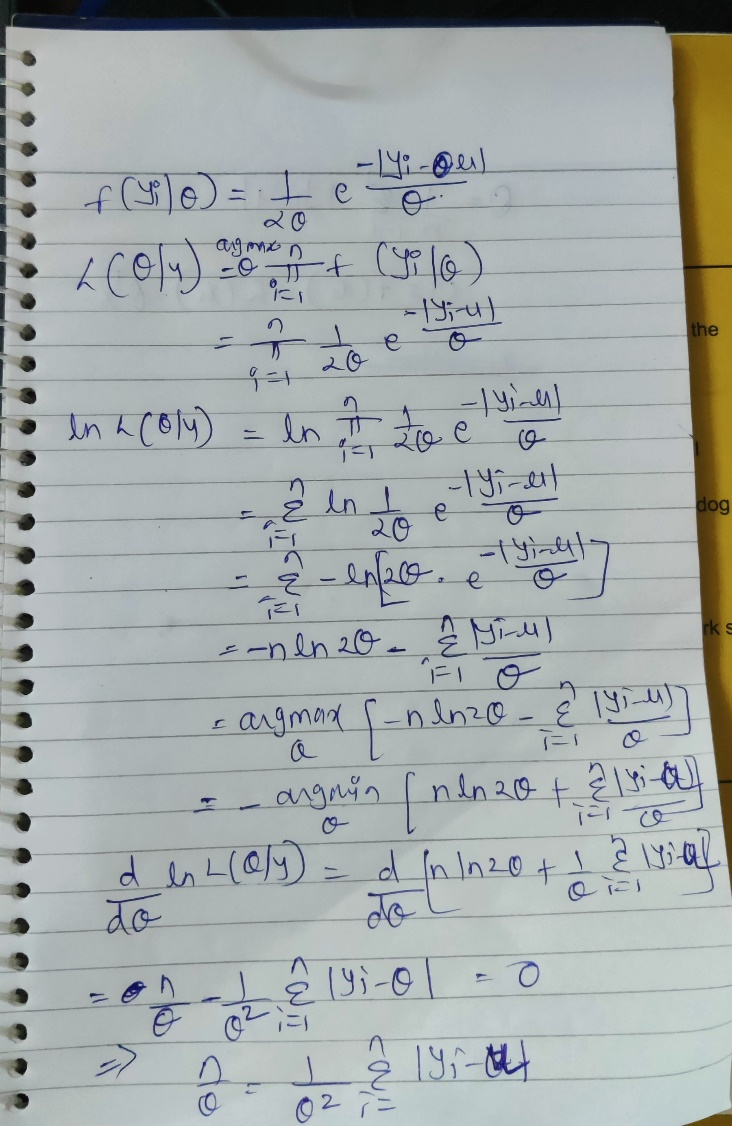
valued of the form  , where *f(xi)* is the true function and *ei* is a random

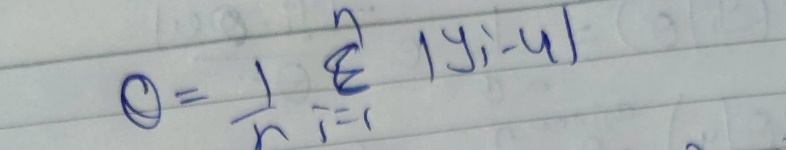
variable representing laplacian noise with PDF given by



Implementing a linear regression model of the form, and,

find the maximum likelihood estimator of *.* Comment on the loss function. [4+1 Marks]





Comment on Loss function: Instead of MSE, MAE is the maximum likelihood hypothesis. So MAE is appropriate for the loss function.

1. Consider a result prediction system where student’s efforts are encoded as percent of

time a student has spent studying out of total available time.

* The input X is having just one feature representing the student’s efforts having only four discrete values (25%, 50%, 75%, and 100%)
* The output Y is having 3 classes (First class, Second class, Fail)
* The priors for each class are: P(Y = First Class) = 0.5, P(Y = Second class) = 0.3, and P(Y = Fail) = 0.2.
* Based on the past data, the estimated the class-conditional probability

P(X| Y) are shown in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| Student’s efforts | p(x|y=fail) | p(x|y=second class) | p(x|y=first class) |
| 25 | 0.7 | 0.4 | 0.1 |
| 50 | 0.2 | 0.3 | 0.1 |
| 75 | 0.1 | 0.2 | 0.3 |
| 100 | 0 | 0.1 | 0.7 |

Consider a following loss function  and y is true class label:



Consider modified Naïve Bayes hypothesis function:



Use this modified hypothesis function to classify each of the examples in the given table. [5 Marks]

Solution:

|  |  |  |  |
| --- | --- | --- | --- |
|  | p(y|x) |  |  |
|  | Fail | second class | first class |
| 25 | 0.35 | 0.12 | 0.02 |
| 50 | 0.1 | 0.09 | 0.02 |
| 75 | 0.05 | 0.06 | 0.06 |
| 100 | 0 | 0.03 | 0.14 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | L(y-hat,y) \*p(y/x) | L(y-hat,y) \*p(y/x) | | L(y-hat,y) \*p(y/x) |  |  |
|  | Fail | second class | first class | | Highest value |  |
| 25 | 0.35 | 0.24 | 0.08 | | 0.35 | fail |
| 50 | 0.1 | 0.18 | 0.08 | | 0.18 | second class |
| 75 | 0.05 | 0.12 | 0.24 | | 0.24 | first class |
| 100 | 0 | 0.06 | 0.56 | | 0.56 | first class |

1. If we modify the loss function of the linear regression model as follows:



Where *w(i)* is the weight assigned to each training example. Derive the equation to

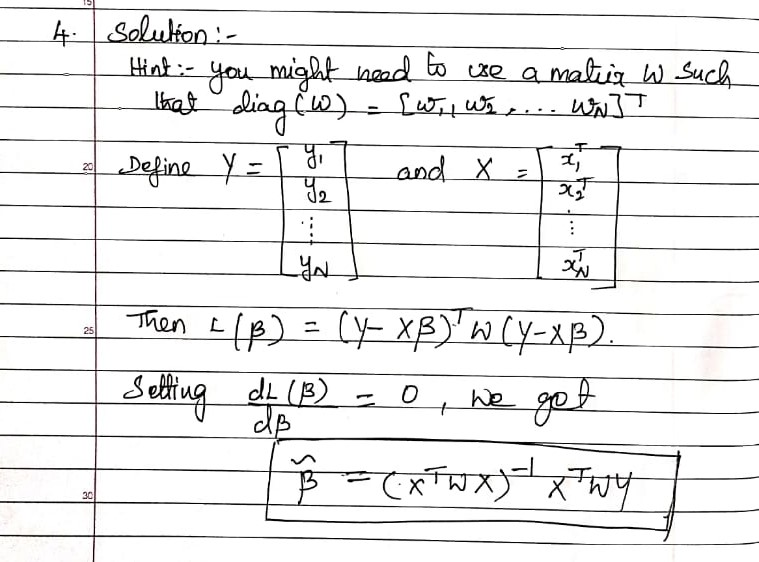
find the value of  with this modified loss function. Suppose, we estimate the value

of *w(i)* inversely proportional to the variance of the residuals, comment in **no more**

**than 20 words** when you prefer to use this kind of modified loss function.

[3+2=5 Marks]

Solution:



Comment: Robust against outliers. Outliers will have higher variance of the residuals resulting into lower weight.

1. Fit a logistic regression. Find the updated weights after the 3 iterations of modified

Gradient Descent algorithm where gradient update happens after every training

example using a learning rate of 0.5 and initial weights (W0, W1, W2) = (1, 1, 1) for the

following data with the logistic regression output given by



Assume the results obtained after 3 iterations is the final weights. Using this construct

the confusion matrix for given below training data. [4+2=6 Marks]

|  |  |  |
| --- | --- | --- |
| Input X1 | Input X2 | Output Label |
| 2 | 0 | 0 |
| 0 | 2 | 0 |
| 0 | -2 | 0 |
| -2 | 0 | 0 |
| 0 | 1 | 1 |
| 0 | -1 | 1 |

Solution:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | Wi-LR\*[Y-Pred – Y ]\* Xi | | |
| X1 | X2 | y | Y-Pred = h(X) | W0 | w1 | w2 |
| 2 | 0 | 0 | 0.05 | 0.975 | 0.95 | 1 |
| 0 | 2 | 0 | 0.95 | 0.5 | 0.95 | 0.05 |
| 0 | -2 | 0 | 0.27 | 0.365 | 0.95 | 0.32 |
| -2 | 0 | 0 | 0.05 |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| w0 | w1 | | w2 |  |  |  |  |  |  |
| 0.365 | 0.95 | 0.32 | |  |  |  |  | True Class | |
| X1 | X2 | y | | Y-Pred = h(X) | Y | Confusion Matrix | | Y=0 | Y=1 |
| 2 | 0 | 0 | | 0.03 | 0 | Predicted Class | Y=0 | 3 | 0 |
| 0 | 2 | 0 | | 0.73 | 1 | Y=1 | 1 | 2 |
| 0 | -2 | 0 | | 0.43 | 0 |  |  |  |  |
| -2 | 0 | 0 | | 0.03 | 0 |  |  |  |  |
| 0 | 1 | 1 | | 0.66 | 1 |  |  |  |  |
| 0 | -1 | 1 | | 0.51 | 1 |  |  |  |  |

1. Consider the following set of training examples:

|  |  |  |  |
| --- | --- | --- | --- |
| Instance | Classification | A1 | A2 |
| 1 | + | T | T |
| 2 | + | T | T |
| 3 | - | T | F |
| 4 | + | F | F |
| 5 | - | F | T |
| 6 | - | F | T |

What is the information gain of A2 relative to these training examples? Provide the equation for calculating the information gain as well as intermediate results. [3 Marks]

