(EC-2 Makeup)

Q.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 $y=e^{(\alpha * x)}$. Using linear regression,

(a) Find the best value of α.

(b) Find the optimal total sum of square error.

Solution:

1. Solution :-

a) Find the best value of α

$\ln(y\_i) = \alpha X\_i$

$J(\alpha) = (\alpha-2)^2 + (2\alpha-4)^2 + (3\alpha-6.3)^2 + (4\alpha-9.2)^2$

Taking derivative of $J(\alpha)$ w.r.t α and equating to 0.

$(\alpha-2) + 2(2\alpha-4) + 3(3\alpha-6.3) + 4(4\alpha-9.2) = 0$

or

$\alpha = 65.7/30$

$= 2.2.$

b) Find the optimal total sum of square error.

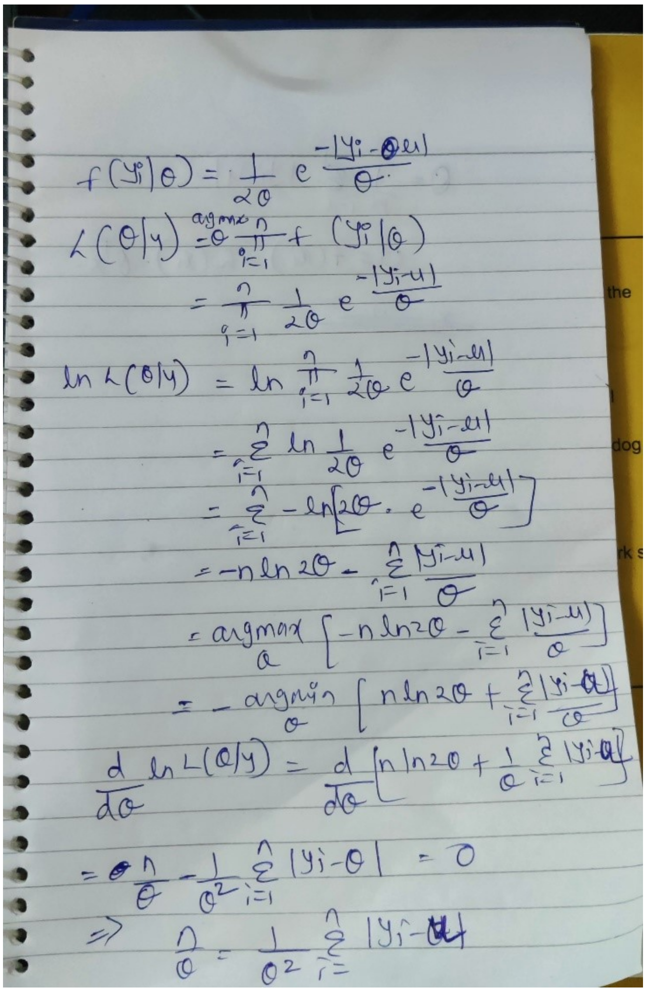
$(\exp(2.2) - \exp(2))^2 + (\exp(4.4) - \exp(4))^2$

$+ (\exp(6.6) - \exp(6.3))^2 + (\exp(8.8) - \exp(9.2))^2$

Q.2 Consider inputs x_i which are real valued attributes and the outputs y_i which are real valued of the form $y_i = f(x_i) + e_i$, where $f(x_i)$ is the true function and e_i is a random variable representing laplacian noise with PDF given by

$$f\left(y_i/\theta\right) = \frac{1}{2\theta} * e^{\frac{-|y_i-\mu|}{\theta}}$$

Implementing a linear regression model of the form $h(x_i) = \sum_{i=0}^{n} \theta_i x_i$, and $\mu = h(x_i)$, find the maximum likelihood estimator of (. Comment on the loss function. [4+1 Marks]



$$f(y_i|\theta) = \frac{1}{2\theta} e^{\frac{-|y_i-\mu|}{\theta}}$$

$$L(\theta|y) \overset{\arg\max_\theta}{=} \prod_{i=1}^{n} f(y_i|\theta)$$

$$= \prod_{i=1}^{n} \frac{1}{2\theta} e^{\frac{-|y_i-\mu|}{\theta}}$$

$$\ln L(\theta|y) = \ln \prod_{i=1}^{n} \frac{1}{2\theta} e^{\frac{-|y_i-\mu|}{\theta}}$$

$$= \sum_{i=1}^{n} \ln \frac{1}{2\theta} e^{\frac{-|y_i-\mu|}{\theta}}$$

$$= \sum_{i=1}^{n} -\ln\left[2\theta \cdot e^{\frac{-|y_i-\mu|}{\theta}}\right]$$

$$= -n \ln 2\theta - \sum_{i=1}^{n} \frac{|y_i-\mu|}{\theta}$$

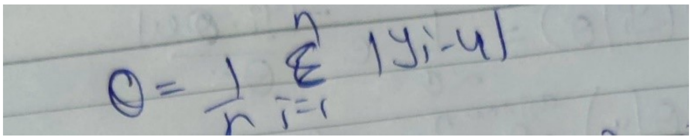
$$= \arg\max_\theta \left[-n \ln 2\theta - \sum_{i=1}^{n} \frac{|y_i-\mu|}{\theta}\right]$$

$$= -\arg\min_\theta \left[n \ln 2\theta + \sum_{i=1}^{n} \frac{|y_i-\theta|}{\theta}\right]$$

$$\frac{d}{d\theta} \ln L(\theta/y) = \frac{d}{d\theta}\left[n \ln 2\theta + \frac{1}{\theta} \sum_{i=1}^{n} |y_i-\theta|\right]$$

$$= \frac{n}{\theta} - \frac{1}{\theta^2} \sum_{i=1}^{n} |y_i - \theta| = 0$$

$$\Rightarrow \frac{n}{\theta} = \frac{1}{\theta^2} \sum_{i=} |y_i - \theta|$$



$$\theta = \frac{1}{n} \sum_{i=1}^{n} |y_i - \mu|$$

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

Q.3 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 $\iota(\hat{y}, y)$ *where* $\hat{y} =$ *predicted class label* and y is true class label:

$$\iota(\hat{y}, y) = \begin{cases} 0 & \hat{y} = y \\ 1 & \hat{y} = Fail \ and \ \hat{y} \neq y \\ 2 & \hat{y} = \text{Sec}\,ond \ class \ and \ \hat{y} \neq y \\ 4 & \hat{y} = First \ class \ and \ \hat{y} \neq y \end{cases}$$

Consider modified Naïve Bayes hypothesis function:

$$\hat{Y} \leftarrow \underset{y_k}{\text{argmax}}\, l(y, \hat{y}) P(Y = y_k) \Pi_i P(X_i | Y = y_k)$$

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

10 0	0	0.06	0.56	0.56	first class

Q.4 If we modify the loss function of the linear regression model as follows:

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{n} w^{(i)} \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right)^2$$

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:

4. Solution :-

Hint :- you might need to use a matrix W such that $diag(W) = [w_1, w_2, \ldots w_N]^T$

Define $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$ and $X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}$

Then $L(\beta) = (Y - X\beta)^T W (Y - X\beta)$.

Setting $\frac{dL(\beta)}{d\beta} = 0$, we got

$$\boxed{\hat{\beta} = (X^T W X)^{-1} X^T W Y}$$

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

Q.5 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 (W_0, W_1, W_2) = (1, 1, 1) for the following data with the logistic regression output given by

$$\frac{1}{1+e^{\left(-W_0+W_1X_1^2-W_2X_2\right)}}$$

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		
X1	X2	y	Y-Pred = h(X)	Y
2	0	0	0.03	0
0	2	0	0.73	1
0	-2	0	0.43	0
-2	0	0	0.03	0
0	1	1	0.66	1
0	-1	1	0.51	1

Confusion Matrix		True Class Y=0	True Class Y=1
Predicted Class	Y=0	3	0
Predicted Class	Y=1	1	2

Q.6 Consider the following set of training examples:

Instance	Classification	A_1	A_2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

What is the information gain of A_2 relative to these training examples? Provide the equation for calculating the information gain as well as intermediate results. [3 Marks]

6. Solution :-

Entropy $E(S) = E([3+, 3-])$

$$= -(3/6)\log_2(3/6) - (3/6)\log_2(3/6)$$

$\boxed{\text{Entropy } E(S) = 1}$

$$\text{Gain}(S, a_2) = E(S) - (4/6)E(T) - (2/6)E(F)$$

$$= 1 - (4/6) - (2/6) \approx 0$$

$E(T) = E([2+, 2-]) = 1$

$E(F) = E([1+, 1-]) = 1$