Date :- $16^{th}$ October, 2022

33

# Machine Learning (Day-4) :-

## Agenda :-

1) Logistic Regression
2) Performance Matrix
   a) Confusion Matrix
   b) Accuracy
   c) Precision
   d) Recall
   e) F - Beta Score.

## 1) Logistic Regression :-

⇒ The logistic Regression technique involves the dependent Variable, which can be represented in the binary (0 or 1, true or false, Yes or No) Values, Which means that the outcome could only be in either one form of two. For Example, it can be utilized when we need to find the probality of a Successful or Fail Event.

Note :- Logistic Regression is used when the dependent Variable (Target) is categoricals.

### *) Logistic Regression Model :-

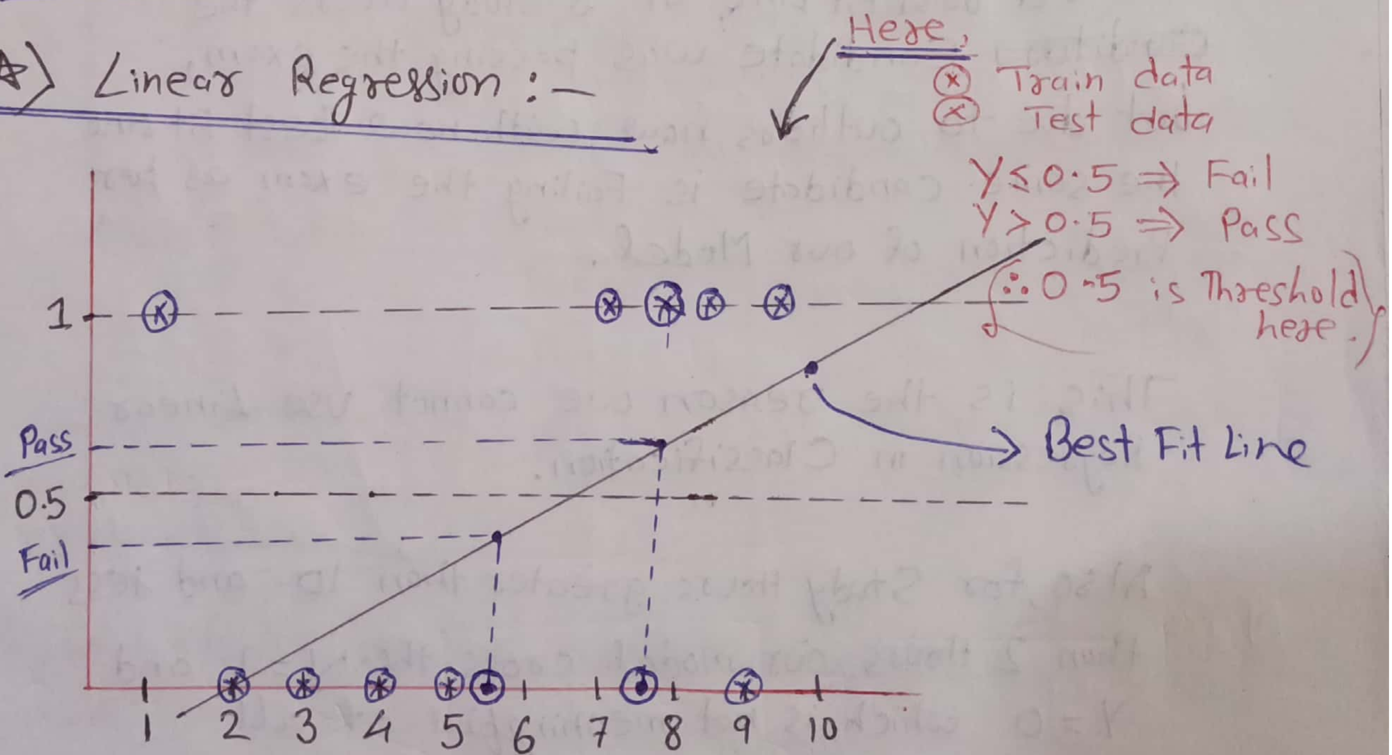

*) Inputs : Study Hours, Play Hours || Weights : $\theta_1$, $\theta_2$ ||
outputs : Pass, Fail.

∴ Logistic Regression is used when the dependent Variable is Target is categorical.

## UPSC Exam Dataset :-

| Study Hours | Output (Pass/Fail) | |
|---|---|---|
| 1 | Pass | ← Outlier |
| 2 | Fail | |
| 3 | Fail | |
| 4 | Fail | |
| 5 | Fail | |
| 6 | Pass | |
| 7 | Pass | |
| 8 | Pass | |
| 9 | Fail | ← Outlier. |
| 9 | Pass | |
| 10 | Pass | |

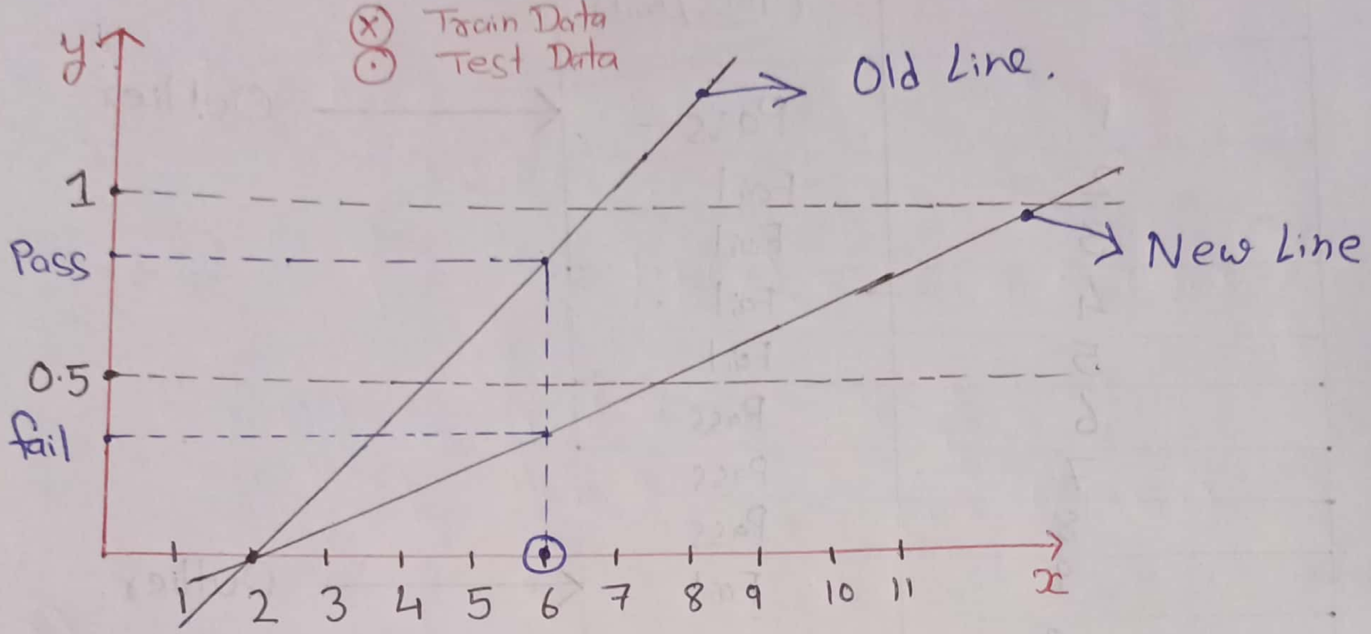
## ★) Linear Regression :-




★

P.T.O

4) Due to Outliers the above best fit line will change. lets observe the changes.




⇒ From above diagram it is clearly visible that with old best Fit Line at 6 Study Hours the ~~Conditure~~ Candidate was passing the exam, but due to outliers now with new best Fit Line the same candidate is Failing the exam as per Prediction of our Model.

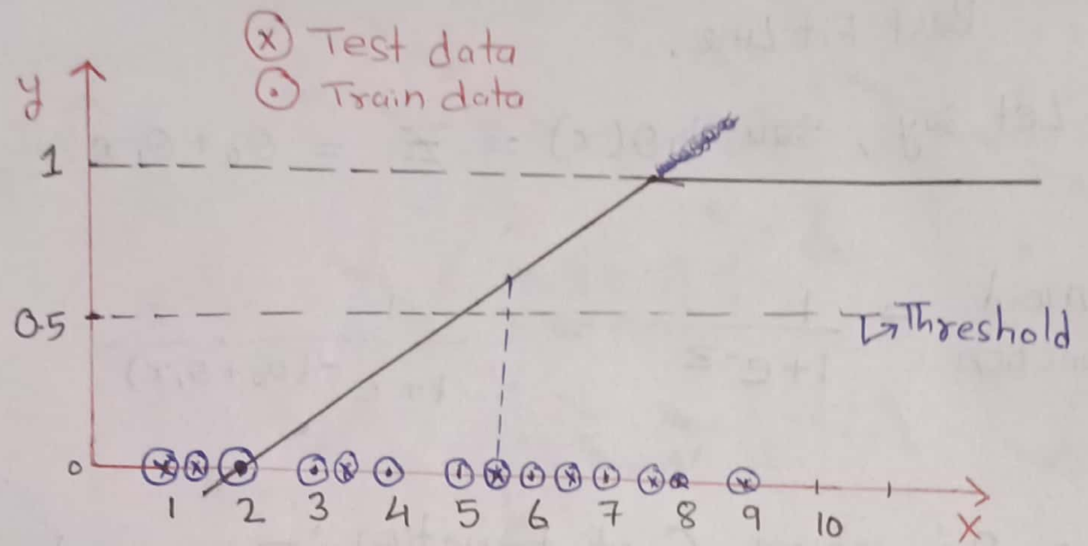
This is the reason we cannot use Linear Regression in Classification.

Also, for Study Hours greater than 10 and less than 2 Hours our model cross the y = 1 and Y = 0 which is not meaningful at all.

This is also an issue with Linear Regression when used for Classification Problem.

P.T.O

So, to Avoid this what we can do is make Best Fit Line Parallel to X-axis when it tries to cross. Y=1 or Y=0 Lines. This is the Logistic Regression Model.




| Test data | Model Prediction |
|---|---|
| 1.5 | Fail |
| 3.6 | Fail |
| 5.4 | Pass |
| 6.7 | Pass |
| 7.9 | Pass |

* To get this type of output that ranges from 0 to 1 we use Sigmoid Activation function. This Function will Squash the Line Parallel to X-axis when it tries to cross y=1 and y=0 Line.

P.T.O

**Step 1 :-** Best Fit Line

$$h_\theta(x) = \theta_0 + \theta_1 x$$

**Step 2 :-** Apply Sigmoid Activation function to Best Fit Line.

Let say, ~~Sa~~ $h_\theta(x) = Z = \theta_0 + \theta_1 x$

$\therefore$ Sigmoid Function $= \dfrac{1}{1+e^{-Z}} = \dfrac{1}{1+e^{-(\theta_0+\theta_1 x)}}$

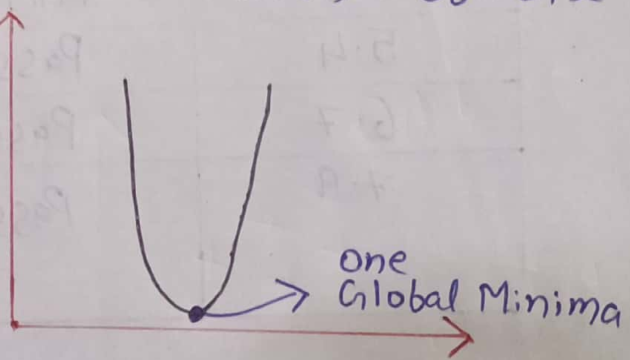
## *) Linear Regression Cost Function :-

**MSE**

$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$$

$\therefore h_\theta(x) = \theta_0 + \theta_1 x$

$\therefore$ This Cost Function has Gradient Descent as convex Function.



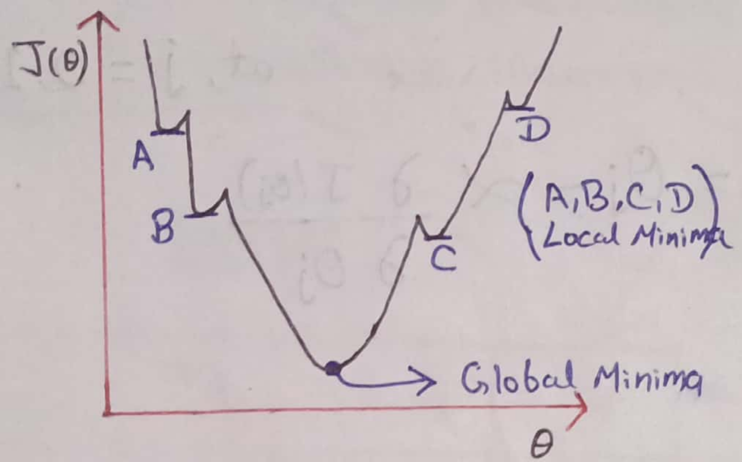

$\therefore$ Now Lets create Logistic Regression Cost Function using above Cost Function.

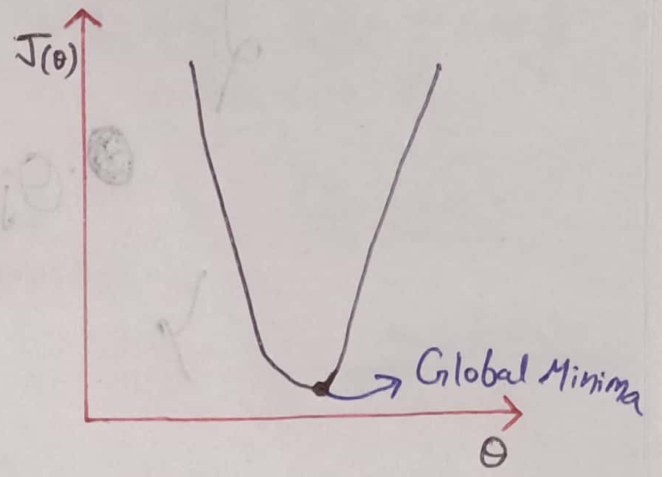
$$h_\theta(x) = Z = \theta_0 + \theta_1 x$$

$$\text{Sigmoid}(Z) = \frac{1}{1+e^{-Z}} = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$$

P.T.O

But, there is one very big issue with this is that it creates a non-convex gradient descent function. Which has lot of Local Minima. Where our gradient-descent algorithm is stuck on the first Local Minima and never reaches global Minima.




A) To Fix this we must change the Cost Function. We use Log-Loss Cost Funktions :-

(Truth Value)

$$\Rightarrow \text{Cost}\left(h_\theta(x)^{(i)}, (y)^{(i)}\right) = \begin{cases} -\log(h_\theta(x)) & , \text{if } y = 1 \\ -\log(1 - h_\theta(x)), & \text{if } y = 0 \end{cases}$$

where, $$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x)}}$$

Simplified Version of Cost Functions.

$$\boxed{\text{Cost}(h_\theta(x), y) = -y \log(h_\theta(x)) - (1-y) \log(1 - h_\theta(x))}$$

↳ This will create a Convex Gradient Descent.

P.T.O

Aim is to minimise the Cost Function $J(\theta_0, \theta_1)$ by Changing $\theta_0, \theta_1$ using Convergence Algorithm.

Repeat Convergence

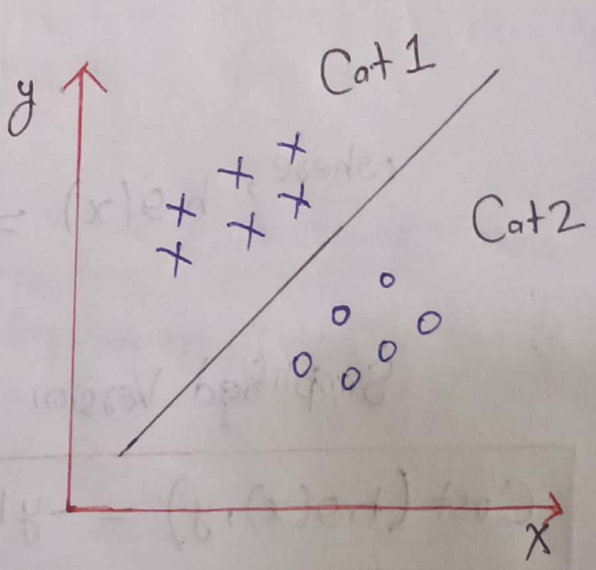
$$\left\{ \theta_j = \theta_j - \alpha \frac{\partial J(\theta_j)}{\partial \theta_j} \right\}$$

at, $j = 0, 1$

## 2) Performance Matrix :-

~~a) Confusion Matrix~~

a) Confusion Matrix
b) Accuracy
c) Precision
d) Recall
e) F- Beta Score.



## a) Confusion Matrix :–

⇒ A confusion Matrix is an extremely useful tools to Observe in which way the model is Wrong (or Right!). It is a matrix that compares the number of predictions for each class that are correct and those that are incorrect.

**Dataset :–**

Model Prediction

| $F_1$ | $F_2$ | o/p | $\hat{y}$ |
|---|---|---|---|
| — | — | 0 | 1 |
| — | — | 1 | 1 |
| — | — | 0 | 0 |
| — | — | 1 | 1 |
| — | — | 1 | 1 |
| — | — | 0 | 1 |
| — | — | 1 | 1 |
| — | — | 1 | 0 |
| — | — | 0 | 0 |
| — | — | 1 | 1 |

($y$) ← Actual Values

| | 1 | 0 |
|---|---|---|
| 1 | TP | FP |
| 0 | FN | TN |

↑ Predicted Value ($\hat{y}$)

### ☆ For Binary Classification :–

← Actual Value ($y$)

| | 1 | 0 |
|---|---|---|
| 1 | 5 | 2 |
| 0 | 1 | 2 |

Truly Positive (5)

Falsely Positive (2)

Falsely Negative (1)

Truly Negative (2)

Predicted Value $\hat{y}$

In a Confusion Matrix, there are 4 Numbers to Pay attention to:–

← Actual Values ($y$)

| | 1 | 0 |
|---|---|---|
| 1 | TP | FP |
| 0 | FN | TN |

↑ Predicted Values ($\hat{y}$)

i) True-Positives (TP):– The Number of Positive observations the model correctly Predicted as Positive.

ii) False-Positive (FP):– The Number of Negative Observations the Model incorrectly Predicted as Positive.

iii) True-Negative (TN):– The Number of Negative Observations the model corretly Predicted as Negative.

iv) False-Negative (FN):– The Number of Positive Observations the model incorrectly Predicted as Negative.

# *) Important Note :-

| | Prediction |
|---|---|
| TP – Truly Positive | Correct |
| FP – Falsely Positive | Incorrect |
| FN – Falsely Negative | Incorrect |
| TN – Truly Negative | Correct |

## b) Accuracy :-

Accuracy is defined as the ratio of correct Predictions to total Numbers of Observations/Predictions.

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

$$\boxed{\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}}$$

*) For Our Confusion Matrix :-

| | 1 | 0 |
|---|---|---|
| 1 | 5 | 2 |
| 0 | 1 | 2 |

$$\text{Accuracy} = \frac{5+2}{5+2+2+1} = \frac{7}{10} = 0.7$$

∴ That is, Our Model is 70% Accuracy.

## *) Limitation of Accuracy :-

⇒ Consider a Binary Classification Data Set with

1000 Observations

900 Pass        100 Fail

∴ Imbalance Dataset

⇒ If we create a dump Model which Predicts only Pass. So for our data set it will predict Correctly for 900 Observations.

So, Accuracy $= \frac{900}{1000} = 0.9$ or 90%

Conclusion :- We cannot decide whether a model is good or Bad by just Considering Accuracy.

P.T.O

c) Precision :– Precision is defined for Pass or fail. i.e Precision of Pass or Precision of Fail.

$$\text{Precision} = \frac{\text{Correctly Predicted Pass or Fail}}{\text{Actual Pass or Fail}}$$

$$(\text{Precision})_{Pass} = \frac{TP}{TP+FP}$$

$$(\text{Precision})_{Fail} = \frac{TN}{TN+FN}$$

Example

Where,

$$(\text{Precision})_{P} = \frac{5}{7}$$

$$(\text{Precision})_{F} = \frac{2}{3}$$

*) The main aim in Precision is to reduce wrong Predictions. (i.e. False Positive and False Negative).

Example :–

i) Spam Prediction :– If our Mail is ham but our model Predicts it as Spam, we may Miss out on important mails, we must reduce this, False Negative.

ii) Diabetes Prediction :– If we have Diabetes, but our Model Predicts that we don't have Diabetes, we will be in quite Danger of getting Seriously ill. We must reduce this, False Negatives.

## d) Recall :- Recall is also defined for Pass or Fail.

$$\text{Recall} = \frac{\text{Correctly Predicted Pass or Fail}}{\text{Total Prediction of Pass or Fail}}$$

$$\boxed{(\text{Recall})_{Pass} = \frac{TP}{TP+FN}}$$

$$\boxed{(\text{Recall})_{Fail} = \frac{TN}{TN+FP}}$$

| | 1 | 0 |
|---|---|---|
| 1 | TP | FP |
| 0 | FN | TN |

### Example

| | 1 | 0 |
|---|---|---|
| 1 | 5 | 2 |
| 0 | 1 | 2 |

$$(\text{Recall})_P = \frac{5}{5+1} = \frac{5}{6}$$

$$(\text{Recall})_F = \frac{2}{2+2} = \frac{2}{4} = \frac{1}{2}$$

### *) Example :-

⇒ Predict if Tommorrow Stock is going to Crash.

| | 1 | 0 |
|---|---|---|
| 1 | TP | FP |
| 0 | FN | TN |

**Model for Consumer**

Aim :- Aim will be to reduce False Negative, So that People can Remove money in Time. FN↓

**Model for Company.**

Aim :- Aim is to reduce False Positive, So that company can, Sell their stock to reduce losses. FP↓

## e) F-Beta Score :-

$\therefore$ F-$\beta \Rightarrow \underset{\text{worst}}{0}$ to $1$ → Optimal

⇒ It is weighted Harmonic Mean of Precision and Recall. $\beta$-Parameter Determines the weight of Recall in the combined Score.

$\beta < 1 \Rightarrow$ More Weight to Precision
$\beta > 1 \Rightarrow$ More Weight to Recall
$\beta = 0 \Rightarrow$ Only Precision
$\beta = \infty \Rightarrow$ Only Recall.

$$\boxed{\text{F-}\beta \text{ score} = \frac{(1+\beta)^2 (\text{Precision} \times \text{Recall})}{\beta^2(\text{Precision}) + \text{Recall}}}$$

i) If we need to give equal important to FP and FN reduction. ($\beta = 1$)

$$\boxed{\text{F-1 Score} = \frac{2 \times P \times R}{P + R}}$$

ii) If FP is more Important than FN. ($\beta = 0.5$)

$$\boxed{\text{F-0.5 score} = \frac{1.25\ P \times R}{0.25\ P + R}}$$

iii) If FN is more Important than FP. ($\beta = 2$)

$$\boxed{\text{F-2 Score} = \frac{5\ P \times R}{4P + R}}$$