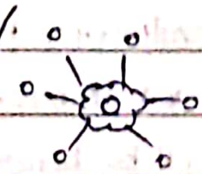
2) Logistic Regression:

- Logistic Regression is a predictive modelling technique. It always involve prediction. It estimate the relationship between a dependent variable (target) and an Independent variable (predictor).
- Dependent variable is a nothing but a variable which you want to a predict and independent variable different feature or values called as independent variable.
- Logistic Regression produce results in a binary format which is used to predict the outcomes of a categorical dependent variable. so the outcome should be descrete/categorical, such as

0 or 1
Yes or NO.
True or false.
High and low



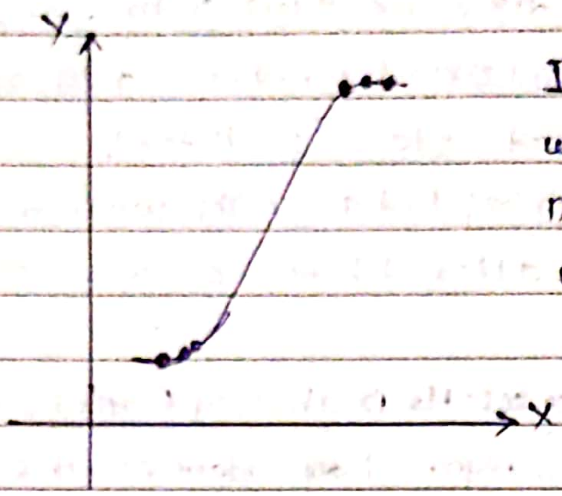
- use sigmoid function/curve to predict the categorical value. Threshold value decides the outcomes (win/lose).

- Linear Regression equation, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_n X_n$.
- Y stands for the dependent variable that need to be predicted.
- β_0 is the Y-intercept, which basically point on the line which touches
- Y axis
- β_1 is the slope of line (slope can be positive or Negative depending on the relationship between the dependent variable and independent variable).
- $X \rightarrow$ independent variable that is used to predict our resultant dependent value.

- sigmoid function

$$P = \frac{1}{1 + e^{-Y}}$$

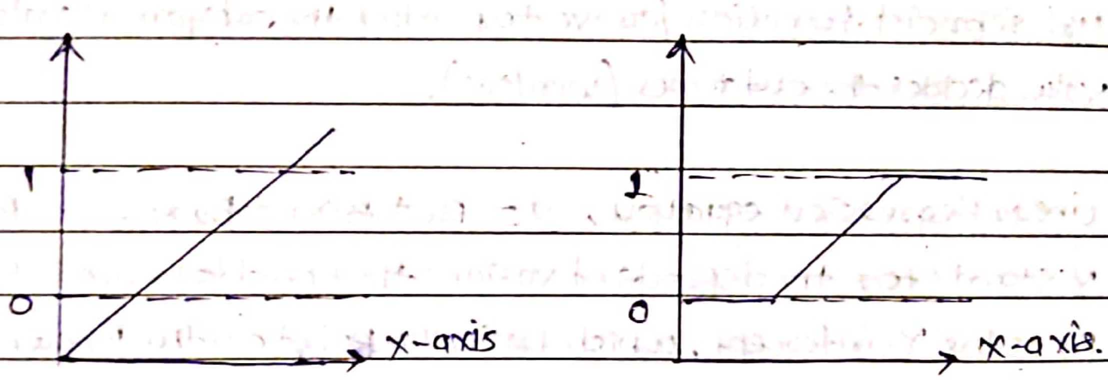
- Apply sigmoid function on the linear regression equation.



If z goes to infinity, Y(predicted) will become 1 and if z goes to negative infinity, Y(predicted) will become zero.

Why Not Linear Regression:

- Linear regression value of y is need to be predict in a range but in Logistic Regression which has two values (binary) either 0 or 1.
- our value of y will be between 0 and 1, the linear line has to be clipped at 0 and 1.



- In Logistic Regression, don't need to be value below 0 and we don't have an value above 1. so our value of y will be between 0 and 1, this is main rule of logistic Regression.

Assumptions of Logistic Regression.

(1) Binary logistic Regression, requires the dependent variable to be binary and ordinal logistic Regression requires dependent variable to be ordinal.

(ii) Logistic Regression Requires the observation to be independent of each other.

(iii) Logistic Regression requires there to be little or no multicollinearity among the independent variable.

(iv) Logistic Regression assume linearity of independent variable and log odds.

(v) Logistic Regression typically requires a large sample size.

Types of Logistic Regression.

- There are three types of logistic Regression.
  (i) Binary Logistic Regression
  (ii) Multinomial Logistic Regression.
  (iii) ordinal logistic Regression.

(i) Binary Logistic Regression:

- It has possibly two outcomes
- Example : yes or NO.

(ii) Multinomial Logistic Regression:

It has three or more nominal categories

Example : cat, Dog, elephant.

(iii). ordinal Logistic Regression:

- It has three or more ordinal categories, ordinal meaning that the categories will be in a order.
- Example : User Rating

- sigmoid function curve convert any value from $-\infty$ to ∞ to desired value, which is a logistic regression.
- suppose, consider we have a data points 0.8 now how can you decide value 0 or 1 now. we have a concept of threshold which is basically decides line.

- Threshold value has indicating the probability of winning or loosing, winning has value 1 and loosing has value 0.
- consider we have a datapoints (0.8).
- If the datapoints value are more than threshold value its gives a resultant as 1, and if datapoints (0.8) is less than threshold value (0.5) then it gives a result as 0 (loosing).
- To calculating or findout value of output uses concept Threshold.
- Logistic Regression equation is derived from straight line equation
- Equation of straight line,

$$Y = C + B1\,X1 + B2\,X2 + \cdots + Mn\,Xn.$$ Range from $-\infty$ to ∞

- Lets try to reduce logistic regression equation from straight line equation

$$Y = C + B1\,X1 + B2\,X2 + \cdots$$ in logistic equation Y can only from 0 to 1.

- Now, get a range of x between 0 and infinity, lets transform Y.

$$\frac{Y}{1-Y} \quad Y=0 \text{ then } 0 \quad Y=1 \text{ then } \infty$$ ← Now range is between 0 to ∞.

- Let us transform it further, to get range between $-\infty$ to ∞.

$$\log \frac{Y}{1-Y} - Y = C + B1\,X1 + B2\,X2 + \cdots$$ final logistic regression Equation.

- To understand the concept of Logistic Regression, it is important to understand the concept of odd Ratio (OR), logit function, sigmoid function or logistic function, and cross Entropy and log loss function

(i) odd Ratio:

- odds ratio in odds in favor of a particular event. It is measure of association between exposure and outcomes.

$$\text{Odd Ratio} = \frac{P}{(1-P)}$$

Where,

$P \rightarrow$ probability of positive events.

probability of success $(P) = 0.8$

Probability of failure $(Q) = 1-P = 0.2$.

- odds are ratio of probability of success and probability of failure.

$$\text{odds(success)} = \frac{P}{Q} = \frac{0.8}{0.2} = 4.$$

$$\text{odds(failure)} = \frac{Q}{P} = \frac{0.2}{0.8} = 0.25.$$

ii) Logit function:

- Logit function is algorithm of odd Ratio (log-odds)
- It takes input value in the range 0 and 1 then transform them to value over the entire real number range.
- Lets take probability, then $\frac{P}{(1-P)}$ is the corresponding odds

$$\text{logit}(P) = \log\frac{(P)}{(1-P)}$$

(iii) logistic function or sigmoid function:

- The inverse of a logit function is called logistic function or sigmoid function.
- It is named as sigmoid function due to its characteristic shape.
- Equation of sigmoid function.

$$\phi = \frac{1}{1+e^{-y}}$$

- sigmoid function takes real number values as input and transform them into values in the range $[0,1]$ with an intercept $\phi(z) = 0.5$.

(iv) cross Entropy or loss Loss:

- cross entropy is commonly used to quantify the difference between two probability distributions, This is used in logistic regression.

cost = (Y actual) * Log(Y predicted) - (1 - Y actual) * log(1 - Y Predicted)

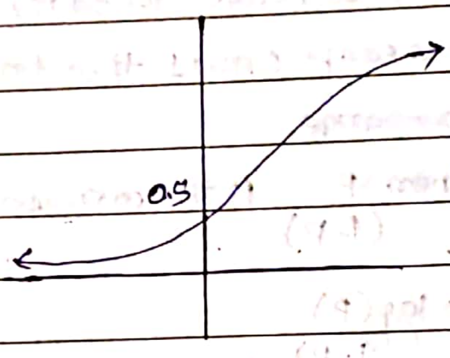
OR.

$\log L(W) = \sum [y^{(i)} \log(\phi(z^{(i)})) + [1 + y^{(i)}] \log(1 - \phi(z^{(i)}))]$

- it is also called as log likelihood function.

$MSE = \frac{1}{n} \sum_{i=1}^{n} (y\ actual - y\ pred)^2$.

→ sigmoid function.



$Y = \frac{1}{1 + e^{-\theta' x}}$.

- cross Entropy or log Loss function.

cost = -(Y actual) * Log(Y Predicted) - (1 - Y ~~predicted~~ actual) * log(1 - Y Predicted)

When y Actual = 1 log (Y predicted) is valid and (1 - Y actual) is zero(0).

When (Y actual) = 0 then (1 - Y actual) is a valid term and log (Y predicted) is zero(0).

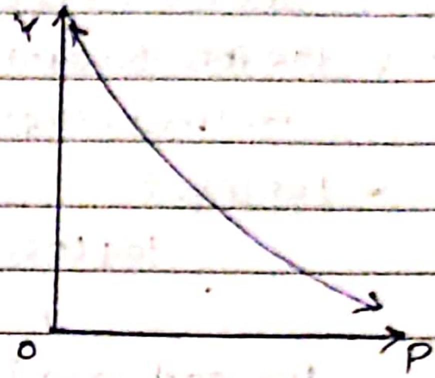
$cost = -\ln(Y\ predicted) \quad Y\ actual = 1$

$\quad\quad\ -\ln(1 - Y\ predicted) \quad Y\ actual = 0.$

If $Y\ actual = 1 \rightarrow -\ln(Y\ predicted)$

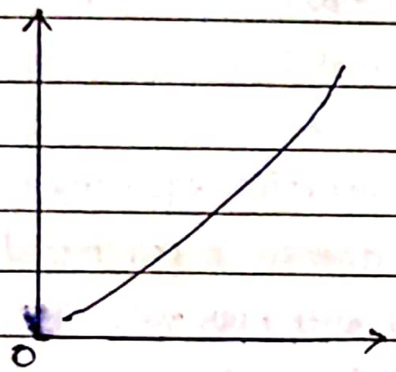
$P = [0\ to\ 1]$ cost function.

$= -\ln(P).$



If $Y\ actual = 0$

$cost = -\ln(1 - Y\ Predicted)$



cost function analyze and wrong prediction.

Y actual	Y predicted	
1	0.95	$\rightarrow MSE = (1-0.95)^2 = 0.025$
1	0.1	$\rightarrow MSE = (1-0.1)^2 = 0.81$

Find out,

$cost = -\ln(Y\ predicted)$

$= -\ln(0.095)$

$= \boxed{0.051}$

$cost = -\ln(Y\ predicted)$

$= -\ln(0.1)$

$\boxed{2.302}$

Scanned with CamScanner

Log Loss functions:

- Estimate against independent variable are generated which reduce the error / loss function.
- log loss function is employed (as opposed to Least square method in linear Regression)
- Log loss:

$$\log loss = -\frac{1}{N}\left(y^i \log(p^i) + (1-y^i)\log(1-p_i)\right)$$

For each row i in dataset,

- Y is outcome (dependent variable) which can be either 0 or 1.
- P is predicted probability outcome by applying logistic Regression

$$P = \frac{e^z}{1+e^z} \quad \Big| \quad z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 - - - - + \beta_n X_n.$$

objective:

- Adjust the estimate in the logistic Regression equation such that the total log loss function over whole dataset is minimized.
- If Y is 1 then function is minimized with high value of P if Y=0 then function is minimized with low value of P.

since log(1) is 0 then this become zero.

$1-1=0$

If Y = 1 then

$$\log loss = -\frac{1}{N}\sum_{i=1}^{N}\left(y^i \log(p^i) + \boxed{(1-y^i)\log(1-p^i)}\right)$$

$$Y * \log(P) = \log(P)$$

$$P \to 1 \text{ where } Y=1 \to 0$$

then become zero

If Y=0 then

$$\log loss = \frac{-1}{N}\sum_{i=0}^{N}\boxed{\left(y^i \log(p^i)\right)} + (1-y^i)\log(1-p^i))$$

$1 * \log(1-P) \to$ if $P=0$ then again where function is zero.

If Y=1 - probability to be high possible., if Y=0 probability to be low possible.

confusion matrix:

- confusion matrix is N x N matrix used for evaluating performance of classification model, where N is the number of target classes.
- matrix compares the actual target values with those predicted by machine learning model.
- for binary classification problem, we could have 2 x 2 matrix

Actual values

		Positive	Negative
predicted values	Positive	TP	FP
	Negative	FN	TN

- columns represent the actual values of target variable.
- Rows represent the predicted values of target variable.

(i) TP (True positive):

- predicted value matches the actual value.
- Actual value was positive and model predict as positive value.

(ii) TN (True Negative):

- predicted value matches actual value.
- Actual value was Negative and model predict as Negative.

(iii) FP (false Positive – Type I Error):

- predicted value was falsely predicted.
- Actual value was Negative but model predict as positive.
- Also known as Type I Error.

(iv) FN (false Negative – Type II Error).

- predicted value was falsely predicted.

- Actual value was positive but model predicted as Negative.
- Also known as type II Error.

(i) Accuracy.

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

Classification Report:

(i) precision:

- precision tell us how many of the correctly predicted cases actu turned out to be positive.

$$Precision = \frac{TP}{TP + FP}.$$

Precision is useful metr incase where false positi is a higher concern than false Negative.

(ii) Recall:

- Recall tell us how many of the actual positive cases we were able to predicted with our model.

$$Recall = \frac{TP}{TP + FN}$$

Recall is a useful metric incas where false Negative trumps fal positive.

- Recall is important in medical cases where it doesn't matter whe we raised a false alarm but the actual positive cases should not go undetected.

- precision is important in music or video recommendations system, e-commerce website etc.

(iii) F1 score :

- We try to increase the precision of our model, the recall goes down, and vice versa, F1 score captures trends in a single value.

$$F1\ score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}.$$

- F1 score is harmonic mean of Precision and Recall.

(iv) Accuracy

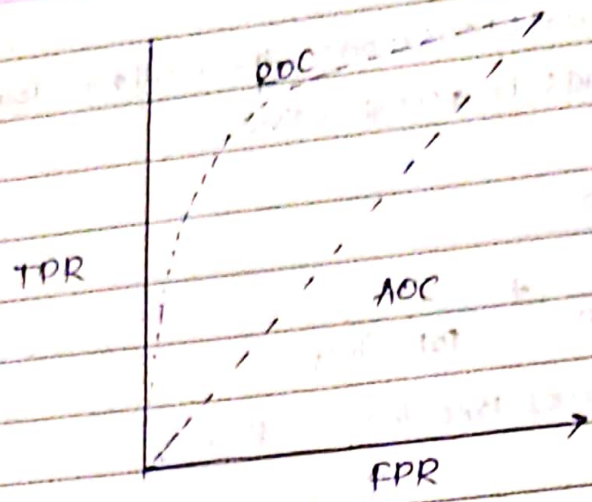
$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}.$$

AUC - ROC CURVE :

- In machine learning, performance measurement is an essential task.
- When it comes to a classification problem, we count on AUC-ROC curve.
- When we need to check or visualize the performance of multiclass classification problem, we use AUC (Area under curve) ROC (Receiver operating characteristics) curve.
- AUROC (Area Under the Receiver operating characteristics).

- AUC - ROC curve is a performance measurement for the classification problem at various threshold settings.
- ROC is a probability curve and AUC represents the degree or measure of separability.
- Higher the AUC, the better model predicting 0s as 0 and 1s as 1s.

- ROC curve is plotted with TPR against the FPR where TPR is on the Y axis and FPR is on X-axis.



(i) TPR (TRue Positive Rate / Recall / sensitivity).

$$TPR/Recall/sensitivity = \frac{TP}{TP + FN}.$$

(ii) specificity

$$specificity = \frac{TN}{TN + FP}.$$

(iii) FPR

$$FPR = 1 - specificity.$$
$$= \frac{FP}{FP + TN}.$$

- AUC near to the 1 which means it has a good measure of separa-bility.
- A poor model has AUC near to the 0 which means it has worst measure of separability.