BASIC MATHS for ML!

Google Developer Student Clubs

Vector and Matrix

CGPA(1-5)
2.62
3.51
4.19
2.35
4.89
2.11

CGPA(1-5)	Technical Skill(0-0.5)
2.62	0.28
3.51	0.26
4.19	0.24
2.35	0.31
4.89	0.12
2.11	0.4

Probability(Basic)

Placed or not
Placed
Not Placed
Not Placed
Placed
Not Placed
Placed

P(A)=n(A)/n(S)

Where S=Sample Set

A=Particular event

n(A)=number of times A Occured

n(S)=Total numbers of outcomes

Conditional Probability

-> Probability of occurring an event knowing another event has already occured

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = P(A \cap B) \,/\, P(A)$$

$$P(A|B) = P(A \cap B) \,/\, P(B)$$

$$P(A \cap B) = P(B|A)\, P(A) = P(A|B)\, P(B)$$

Example of Conditional Probability

How many students got placed who have cgpa>4 ?

A: Students got placed

B: Student got CGPA>4

P(A|B)=?

StdID	CGPA(1-5)	Technical Skill(0-0.5)	Actually Placement
sid001	2.62	0.28	Placed
sid002	3.51	0.26	Placed
sid003	4.19	0.24	Not Placed
sid004	2.35	0.31	Placed
sid005	4.89	0.12	Not Placed
sid006	2.11	0.4	Placed
sid007	2.9	0.25	Not Placed
sid008	3.25	0.29	Placed
sid009	3.9	0.15	Not Placed
sid010	4.02	0.2	Not Placed

Bayes' Theorem

Bayes' Theorem provides a way to revise existing predictions or theories (update probabilities) given new or additional evidence.

$$P\left(B_J \mid A\right) = \frac{P\left(A \mid B_J\right) P\left(B_J\right)}{\sum_{i=1}^{n} P\left(A \mid B_i\right) P\left(B_i\right)}$$

Example of Bayes' Theorem

Now,my algorithm finds a wrong answer while checking the final sem ans sheet, now it tries to predict the probability that the ans is given by std2.

A: The wrong answer is given by std2

B: The answer is wrong

Result of CA

Std 1

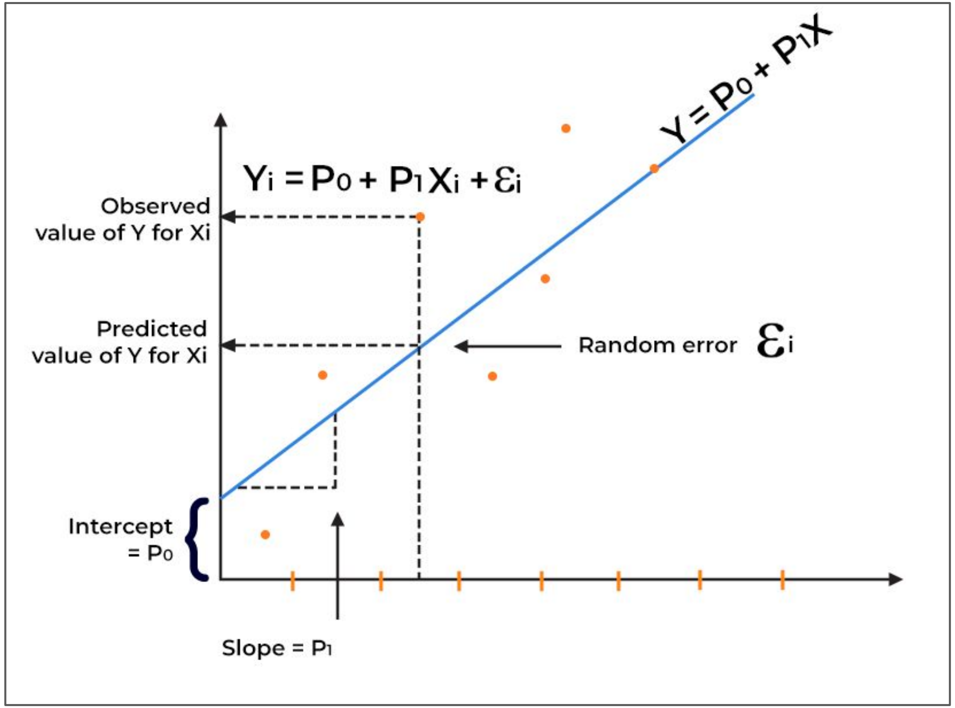
Correct:4
Wrong: 2

Std 2

Correct:5
Wrong:1

Linear Regression

- Linear regression analysis is used to predict the value of a variable based on the value of another variable.
- Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values.



How to find accuracy of Linear Regression?

R2 Score

Formula

$$R^2 = 1 - \frac{RSS}{TSS}$$

R^2 = coefficient of determination
RSS = sum of squares of residuals
TSS = total sum of squares

$$RSS = \Sigma\left(y_i - \widehat{y_i}\right)^2$$

Where: y_i is the actual value and, $\widehat{y_i}$ is the predicted value.

$$TSS = \Sigma\left(y_i - \overline{y}\right)^2$$

Where: y_i is the actual value and $\overline{y}$ is the mean value of the variable/feature

Error Calculation

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

Logistic Regression

CGPA(1-5)	Technical Skill(0-0.5)	Placed or not	Logistic Function value
2.62	0.28	Placed	0.05215356308
3.51	0.26	Not Placed	0.02253263946
4.19	0.24	Not Placed	0.01177420602
2.35	0.31	Placed	0.06537533343
4.89	0.12	Not Placed	0.00662669725
2.11	0.4	Placed	0.07516010948

Equation:

$$f(x) = \frac{1}{1+e^{-(x)}}$$

Confusion Matrix

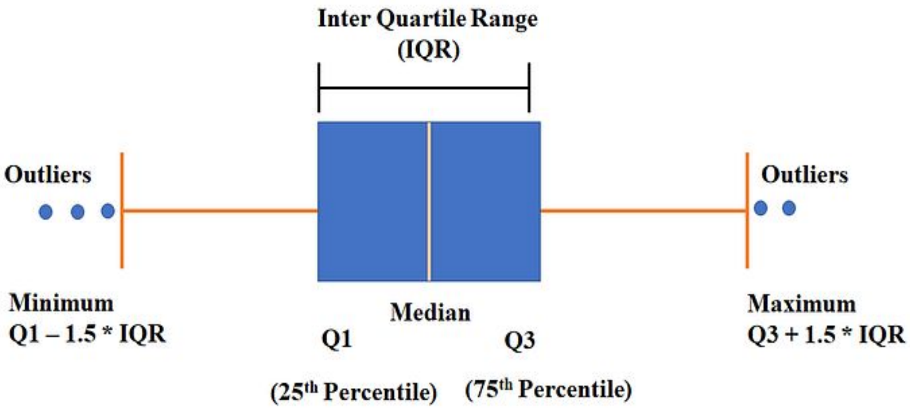
StdID	CGPA(1-5)	Technical Skill(0-0.5)	Actually Placement	Logistic Function value	Predicted Placement
sid001	2.62	0.28	Placed	0.05215356308	Placed
sid002	3.51	0.26	Placed	0.02253263946	Not Placed
sid003	4.19	0.24	Not Placed	0.01177420602	Not Placed
sid004	2.35	0.31	Placed	0.06537533343	Placed
sid005	4.89	0.12	Not Placed	0.00662669725	Not Placed
sid006	2.11	0.4	Placed	0.07516010948	Placed
sid007	2.9	0.25	Not Placed	0.0410912782	Placed
sid008	3.25	0.29	Placed	0.02819528797	Not Placed
sid009	3.9	0.15	Not Placed	0.01712403332	Not Placed
sid010	4.02	0.2	Not Placed	0.014485724	Not Placed

With threshold of :0.04

Outliers and Box-Plot

Let, Std marks:

10,3,15,16,16,13,24



Outliers and Box-Plot

Let, Std CA marks:

10,3,15,16,16,13,24

Re-arranged:

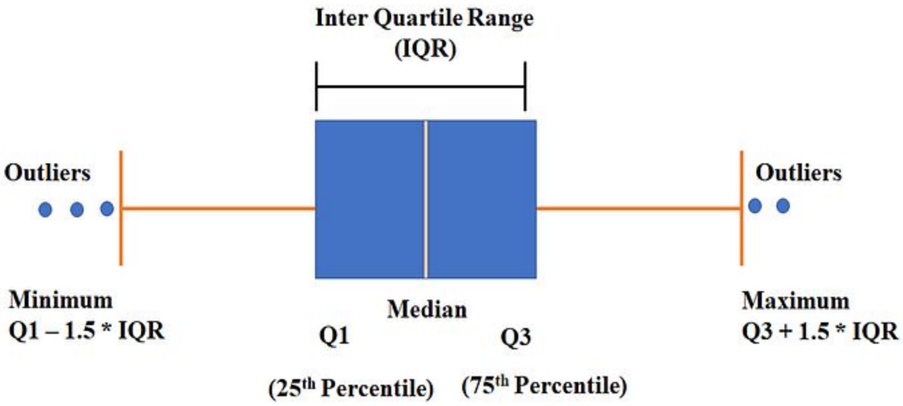
2,11,13,15,16,16,24

Middle Value(median)

Mean=(2+11+13+15+16+16+24)/7

Mode=16

Q1=11.5,Q2=15,Q3=16,IQR=(16-11)=5



Outliers: <(11.5-(1.5*5))
or >(16.0+(1.5*5))

Detailed Derivation

Q1 Calculation:

- ❑ index=0.25×(7-1)=0.25×6=1.5
- ❑ value at index 1=10,value at index 2=13
- ❑ interpolated value=10+0.5×(13-10)=10+0.5×3=10+1.5=11.5

Q3 Calculation:

- ❑ index=0.75×(7-1)=0.75×6=4.5
- ❑ value at index 4=16,value at index 5=16
- ❑ interpolated value=16+0.5×(16-16)=16+0.5×0=16