Day 4

1) Central limit theorem
2) Probability
3) Permutation & Combination

1) Central limit theorem

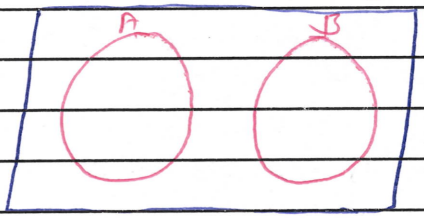
If the population data is normally distributed or not, but if takes sample from population ≥ 30 times & take mean of all sample it would be a normal distribution.

2) Probability.

* Mutual Exclusive Event

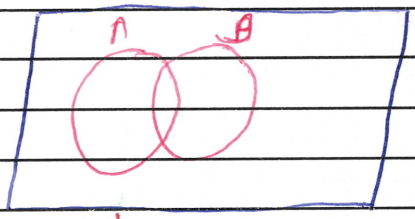
Two Events are mutually exclusive if they cannot occur at the same time

1) Tossing a coin
2) Roling a dice



* Non mutual Exclusive Events

Two Events can occur at the same time.



(Problem statement)

① What is the probability of coin landing on head or tails.

addition rule for mutual exclusive events

$$P(A+B) = P(A) + P(B)$$
$$= \frac{1}{2} + \frac{1}{2} = 1$$

(11) What is the probability of getting 1 or 6 or 3 while rolling a dice?

$$P(1 \text{ or } 6 \text{ or } 3) = P(1) + P(6) + P(3)$$
$$= 1/6 + 1/6 + 1/6$$
$$= 3/6 = 1/2$$

Non mutual exclusive event prob. statements

(1) bag of Marbuls: 10 red, 6 green, 3 (R & G)
When picking randomly from a bag of marbuls what is the probability of choosing a marbul that red or orangegreen?

Addition rule for non mutual exclusive event.
$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

$$\frac{10}{19} + \frac{6}{19} - \frac{3}{19} = \frac{13}{19} \text{ Ans}$$

$$\frac{13}{19} + \frac{9}{19} - \frac{3}{19} = 1 \text{ Ans}$$

* Complement rule
The probability that an event does not occur is equal to one minus the probability that it will occur.

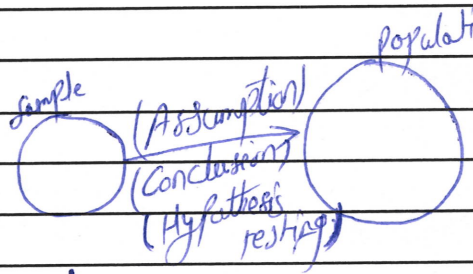
☆ Inferential Statistics ☆ (3 distribution)

① Hypothesis testing
② P-Value
③ Confidence Interval
④ Signifi

- Z test
- t test
- Chi square test
- Anova test (F-test)

- Bernauli
- Binomial
- Power Law

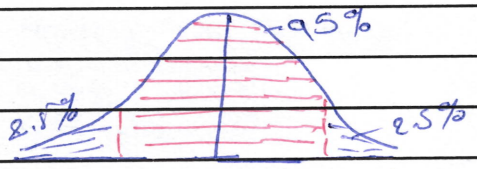
① Hypothesis testing



(Steps of Hypothesis Testing :)

① Find null hypothesis (H_0)
② Try to fail null Hypothesis
③ Using Hypothesis testing

Confidence Interval : (CI)



Significance value

C.I = 95%

S.V. = 1 − .95

= 0.05

S.V = 1 − CI

Point Estimater

☆ Point estimates are estimates of population parameters based on sample data.

$\bar{x}$ → μ (population mean)

Statistics — Paramater

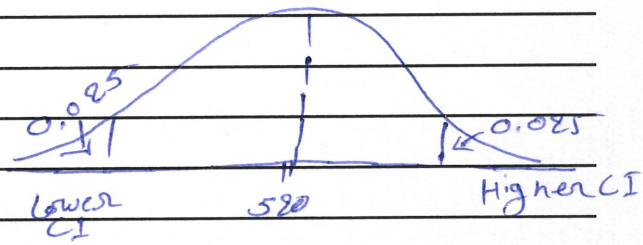
Point Estimate ± margin of error = paramater

(C.I)
Lower Fence = Point estimate - Margin of error
Higher (C.I) Fence = Point estimate + Margin of error

$$\text{Margin of error} = \boxed{Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}} \quad \text{Standard error}$$

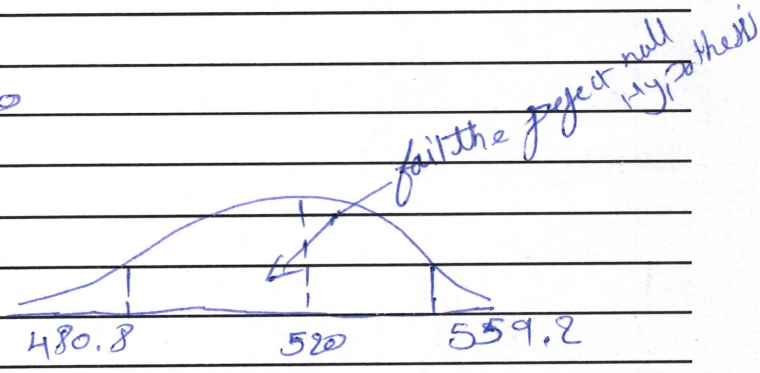
* On the quant test of CAT Exam, a sample of 25 test takers has a mean of 520 with a population standard deviation of 100. Construct a 95% C.I. about the mean?

$n = 25$
$\bar{x} = 520$
$\sigma = 100$
C.I = 95%
SV = 1 - 0.95
= 0.05



Lower C.I. = Point estimate - Margin of error
$= 520 - Z_{0.05/2} \frac{\sigma}{\sqrt{n}}$
$= 520 - Z_{0.025} \frac{100}{\sqrt{25}} \; 20$
$= 520 - 0.975 \times 20$
$= 480.8$

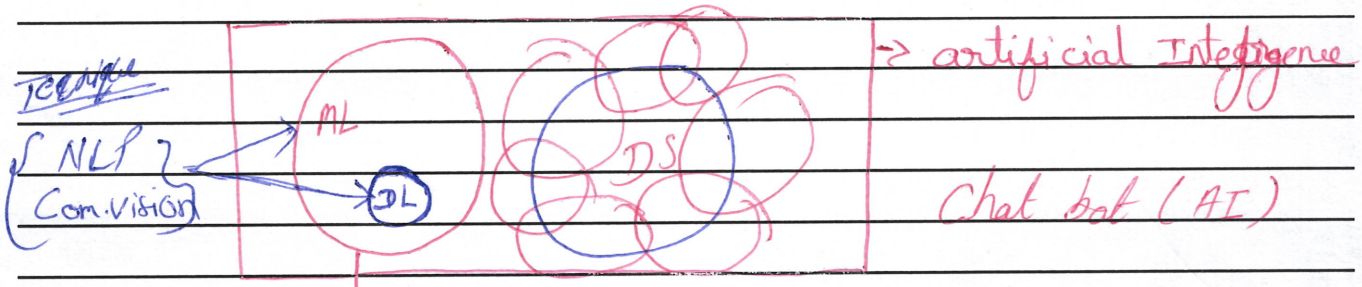
Higher CI = $520 + 1.96 \times 20$
$= 559.2$



Machine Learning Introduction & Simple Linear Reg.

(i) Machine Learning Introduction
(ii) AI Vs ML Vs DL Vs DS
(iii) Simple Linear Regression → Mathematical Intitution

★ AI Vs ML Vs DL Vs DS

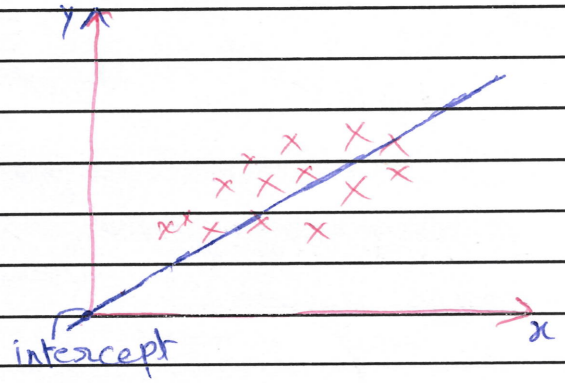


ML product provids Stats tools to analyse, visualize, perform prediction and other task with the help of data.

Machine & Deep Learning

- Supervised
  - Regression
    1. Linear Regression
    2. Polynomial
    3. SVR
    4. Decission Tree
    5. Random forest
    6. Xgboost
    7. KNN
  - Classification
    1. Logistic Regression
    2. SVM
    3. Decision Tree
    4. Random forest
    5. naive Bayes
    6. KNN
- Unsupervised
  - clustering Algorithes
    - DBScan
    - K Means
    - Hierarchical
    - Gaussian

✡ Simple linear regression ✡



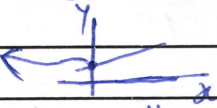
✡ Equation of a straight line

$y = mx + c$

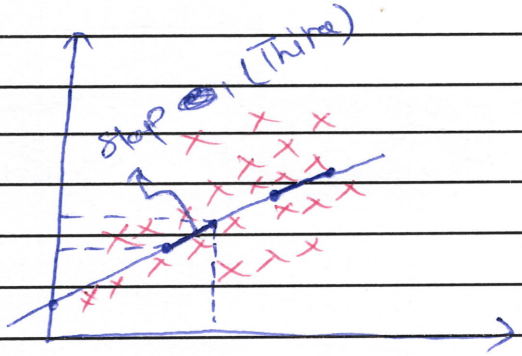
$\hat{y} = \beta_0 + \beta_1 x_1$

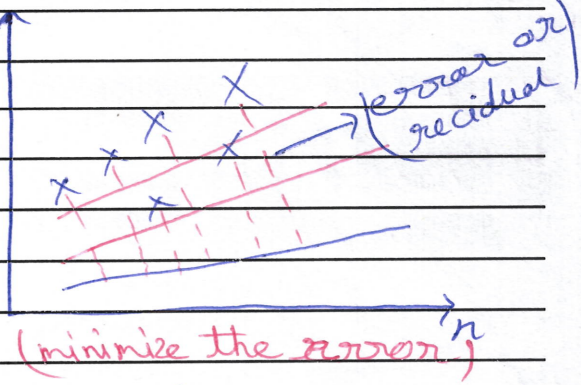
$h_\theta(x) = \theta_0 + \theta_1 x$ → Slop

Intercept (When x is zero(0) where is the line meeting on y axis)



Slop (with the moment in the x axis what is the moment in th y axis)





(minimize the error)

Loss

Cost function

mean Square error

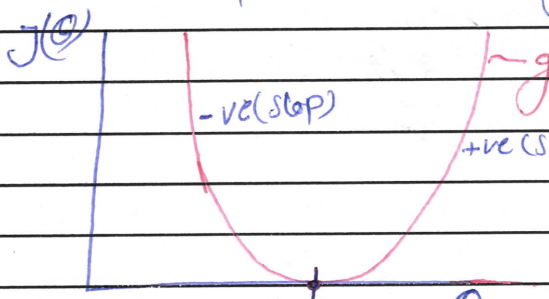
$$J(\theta_0, \theta_1) = \sum_{i=1}^{n} \left(h_\theta(x)^{(i)} - (y)^{(i)}\right)^2$$

(Predicted: $h_\theta(x)^{(i)}$, Actual: $(y)^{(i)}$)

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

★ Gradient Descent curve
★ global minima

(Convergence algorithm)
{Optimize the change of θ_1 value}



repeat until convergence {

Linear rate

$$\theta_j := \theta_j - d \frac{\partial}{\partial \theta_j} J(\theta_j)$$

$$\theta_j = \theta_j - \alpha(-ve)$$
$$= \theta_j + \alpha$$

gradient descent

MSE
↳ Cost function [MAE / MSE / RMSE]

Loss function ⇒ Loss function is a method of evaluating How well your algorithem is modelling your dataset

(i) Why loss function Important?
⇒

(ii) MAE, MSE & RMSE?

(iii) How gradiant decent work?

Q1) What is the Linear regression?
Q2) How we can Calculate errror in linear regression?
Q3) Diff. between loss & cost function?
Q4) " " MAE, MSE & RMSE?
Q5) How gradient decent works? in Linear regression?

Linear Regression → (I) Calculation ($y = mx+c$)
(II) Loss (MSE, MAE, RMSE)
(III) optimize (m,c) - Minimum Loss

Q6) Explain what is the Intercept (term mean)?
Q7) What is the assumption for linear regression?
Q8) How to hypothesis testing use for LR?
Q9) How could ye decided the importance Variable for the multi Variat regression?
Q10) Diff. between R Sqyre & Adjusted R Square?

Ans 1). Linear Regression is a Supervised Machine learning technique to find out numerical value using Independent(x) & dependent(y) Variable.
∴ It has 3 types of regression.
(I) Simple linear regression
(II) multi linear regression
(III) Polynomial regression

Ans 2) We can Calculate the error using Varias type.
(I) MAE → Mean Absolute error
(II) MSE → Mean Squared error
(III) RMSE → Root Mean Squared error
(IV) R2 Score → or Coefficient of determination
(V) Adjusted R2 Score.

Ans 3) Loss function :- It's method of evaluating, how good your algorithem to modelling your data. (It's a techinque to find out, your algorithem working fine or not)

if your loss function performance with high value algorithem is poor if it's perfomance with low value algorithem is good.

Cost function :- It's a Error representation in ML. It's shows how our model is predicting Compared to original given data set (if high cost function less accuracy & less cost function is high accuracy)

Ans4. (i) Mean absolute error (MAE) is show the measure of error between predictive data & real data

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

(ii) Mean Squard error (MSE) is Average of the squared error that is use for as the loss function for least Square error

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$$

(iii) Root mean Squard error (RMSE) this function is work same as MSE it give the return in same unit (pa

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}$$

Ans5. Gradient descent is a algorithem use for minimize Cost function value.

Ans 6. When one axis point is zero (0) where the other axis point is meeting.

Ans 7> Types of linear regression assumptions.

(i) Linear Relationship between input & output
(ii) No Multicolinearity
(iii) Normal Residual
(iv) Homoscedasticity
(v) No Autocorrelation of error

Ans 8>

Ans 9>

9th Oct.

① linear regression algorithm

$SLR = h_\theta(x) = \theta_0 + \theta_1 x_1$

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \text{ --- } \theta_n x_n$

Convergence algorithm

$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x)' - y)^2 \Rightarrow$ Cost function

Loss function Vs Cost function

☆ Performance Metrics ☆

1. Confusion Metrics
2. Accuracy
3. Precision
4. Recall
5. F-Beta Score

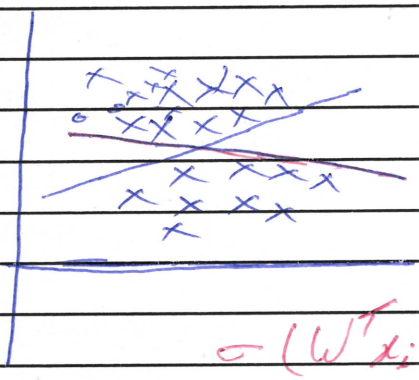
Confusion Metrics

predict \ actual	1	0
1	TP	FP
0	FN	TN

$$acc = \frac{TP+TN}{TP+FP+FN+TN} = \frac{3+1}{3+2+1+1}$$

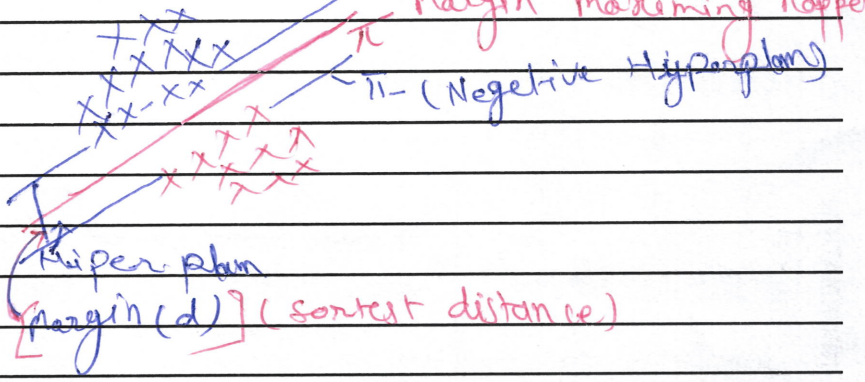
$$= \frac{4}{7} = 57\%$$

SVM (Support Vector Machine)

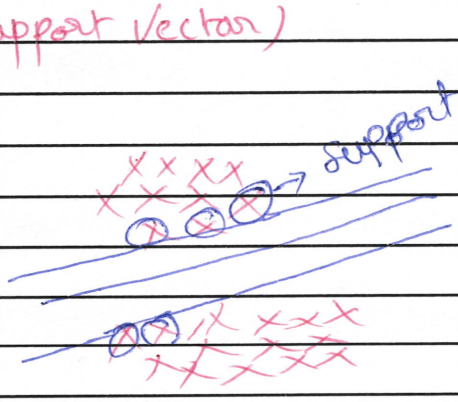


$\sigma(W^T x_i) = P(y=1)$

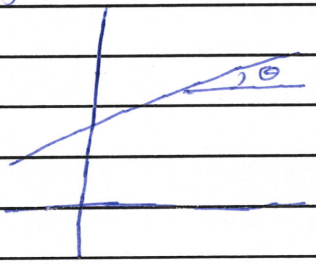
σ sigmoid
T transpos
P Probability



✱ (Concept of Support Vector)



1. Robust to outliers
2. Non Linear data
3. Classification & Regression impliment



$y = mx + b$

$ax + by + c = 0$

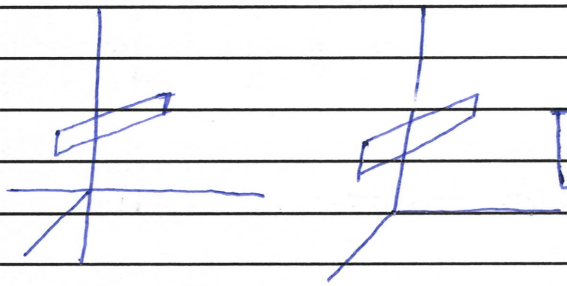
$m = -\frac{a}{b}x - \frac{c}{b}$

$m = -\frac{a}{b} = b = -\frac{c}{b}$

Cf →

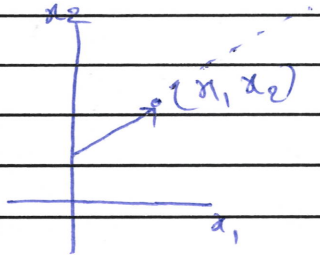
$ax + by + c = 0$

$$w_1 x_1 + w_2 x_2 + w_0 = 0$$



$$w_1 x_1 + w_2 x_2 + w_3 x_3 + w_0 = 0$$

lets assume



dot product

$[w_1, w_2] = w \cdot x = w_1 x_1 + w_2 x_2 + w_3 x_3$

$w \cdot x + w_0 = 0$ * $w_0 = (w\ Nat)$

Vector format

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

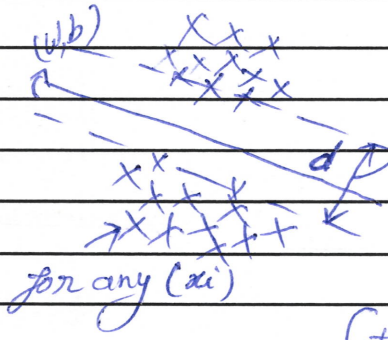
Row Vector Col vector

matrix format $\underline{w_1\ w_2\ w_3\ w_4} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ $w^T x + w_0 = 0$

like $(a \cdot b = a^T b)$

final equation $(w^T x + w_0 = 0)$

(Mathematics of SVM)



(For classify this line we need a line that maximize this distance)

$(w^T x + b = 0)$

(dis margin)

$$\hat{y} = \begin{cases} +1 & \text{if } \vec{w}.\vec{x}_i + b \geq 0 \\ -1 & \text{if } \vec{w}.\vec{x} + b < 0 \end{cases}$$

☆ assumption

① $\pi^+ \quad w^T x + b = 1$

② $\pi^- \quad w^T x + b = -1$