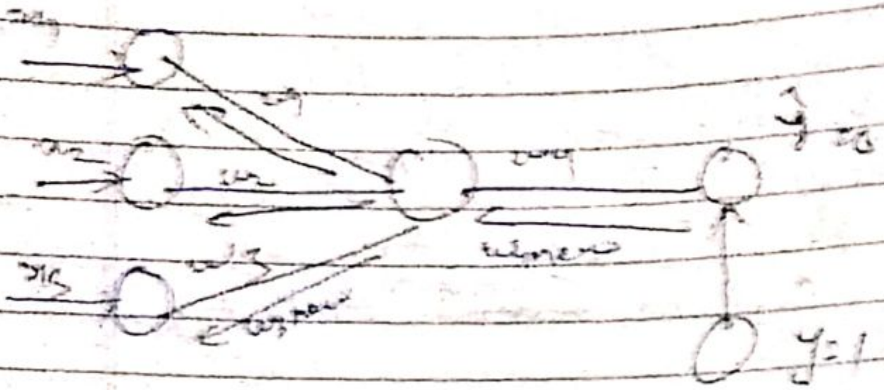
# DEEP Learn

① ANN ② CNN ③ RNN

Play Study Sleep

**Back Propagation** $x_1$ $x_2$ $x_3$




$loss = (y - \hat{y})^2$

$= (1 - 0)^2 = 1$ } this should be reduce ↓

So to reduce lost we do Back propagation
So we need to adjust weight for same

minimise loss
→ optimiser $\eta$, learning rate

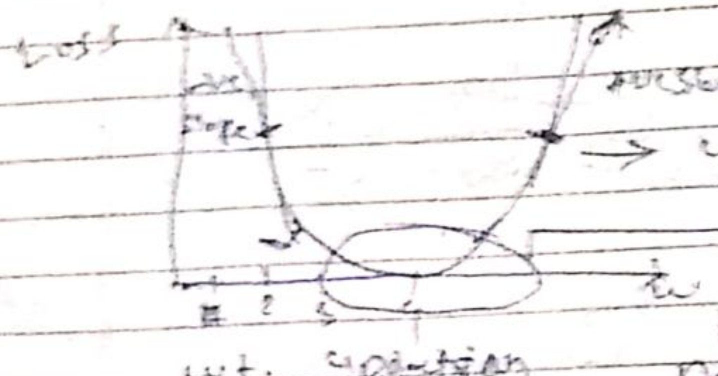
$w_{4new} = w_{4old} - \eta \frac{dL}{dw_4}$

$w_3$, $w_2$ here will calculate as same way

after getting all weights again front propagation will happen

Cost function → For multiple no of neurons
$$= \sum_{i=1}^{n} (y-\hat{y})^2$$
as here we have only o/p 1

Optimizers
* Gradient Decent




→ global minima, our main aim of gradient Decent to achive this area

$W_{new} = W_{old} - \eta \frac{\partial L}{\partial w}$ → Learning Rate should be take, which suit for G.D.

↳ derivitive will give slope of the wt. points.

If slope is -ve then it will add
-11- +ve then it will minus

Unless an until we reach to the global minima updation of wt. will happen with Back Propagation

$loss = (y-\hat{y})^2$, this value must be decreasing so we can achive global minima

~~self cost~~

~~...ary classification~~

ML

(1) Supervised (2) Unsupervised (3) Reinforsmen

Application of supervised

It's process of training algorithm, is like teachers provide training data set to predict flowdone. [feeding to the model & it will predict]

e.g :- Popular Algo

(1) Linear Regre (2) Random Fore (3) SVM

2

Usecase of ~~SVM~~ Supervised

(1) Cortana Voice Recognati

(2) ~~weather~~ Biometric identifie

(3) machine can identify provide i/p patten and identify in futur

(3) ~~Banking~~ Health Care

By Building a Regression model, we can predict Patient Readmission Rates

(4) Retail

By association Rules we can Recommend Customer to Buy something with ~~...~~ Buy together

PAGE NO
DATE / /

Unsupervised Learning

① We have i/p data x and no Corresponding o/p data.
Model make similar things together or make cluster of similar things according to the i/p data
~~Algo~~
Why we called it as Unsupervised.
Algo: teaches own to cluster data together

Algo Eg:- Apriori algo; K-means, Hierarchical Clust

Application:-
Banking: Segment Customers by Behavioural Characteristic
Heathcare:-
Categorize MRI Data by Normal or Abnormal
Retail Sectore:
Recommed product to Customer based on past Purchase

* Reinforcement M/L.
It's determine idea behaviour within specific context to maximise the performance.
Interaction beteen two element, Environment & learning Agent
Model train as if it's get correct decision then it will reward otherwise it will be have Penalty for Same

Scanned by CamScanner

Usecase

Banking:-

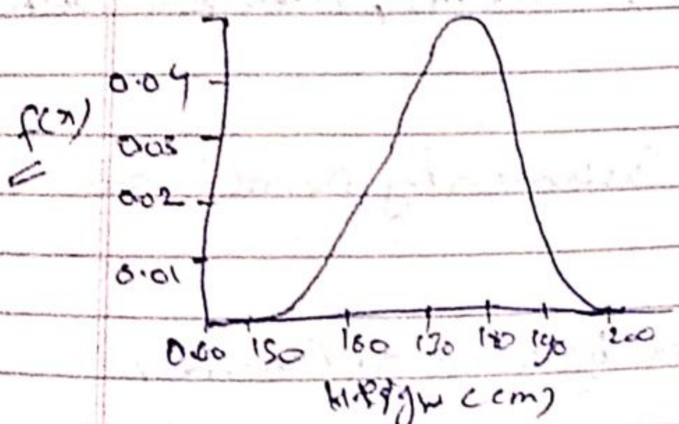
Create Next Best offer for Call Centre

Retail:-

Reduce Excess Stock with Dynamic Pricing

Gaussian Distribution [Normal Distribution]

Its an extremely important continuous probability distribution




f(x) is a probability density function which will give the ht. of (x) etc

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{for } -\infty < x < \infty$$

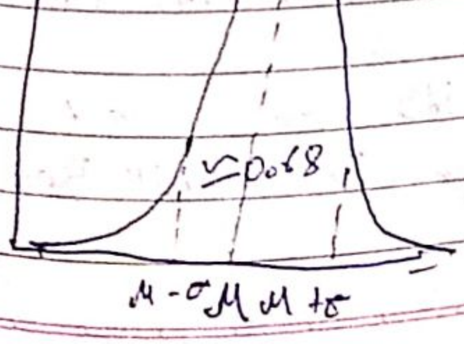
One std deviation area 0.68

Mean

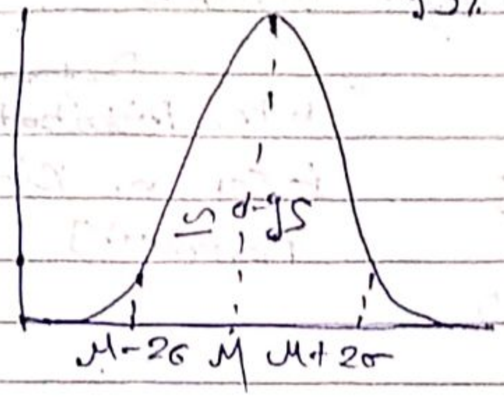
$-\infty < \mu < \infty$

Std. deviation

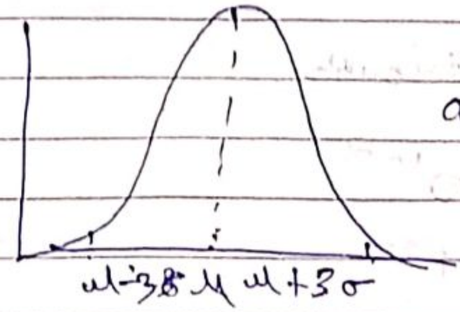
$\sigma > 0$

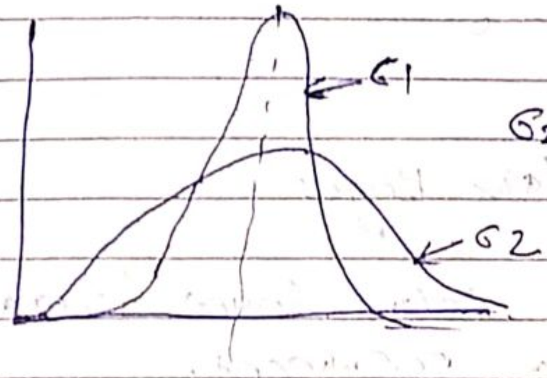


95% area in 2 Std deviation




99% of area in 3rd Std deviation across the mean

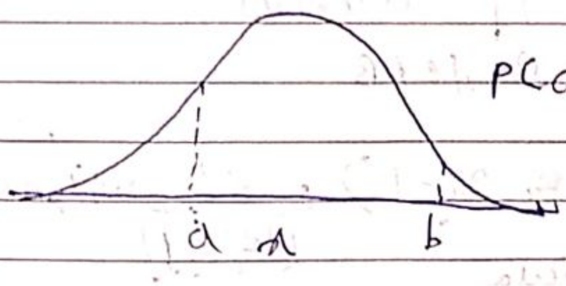



When Std. deviation $\sigma_2 = 2\sigma_1$ is greater than more area in tails & lower peak.

The Std. Normal Distribution is a normal distribution with mean 0 & variance 1.

$$Z \sim N(\underset{\mu}{0}, \underset{\sigma^2}{1})$$




$P(a<x<b)$ We will often need to find area under the Std. normal curve (To find the probability)

PAGE No.
DATE / /

Bernoulli - Naive Bayes Variants

Bernoulli Distribution

$P(success) = p$

$P(Failure) = 1 - p = q$

[ when ~~Distribution~~ feature in data is set are Binary vs Bernoulli ]

Random variable → $X = 1$ [ success ]

$X = 0$ [ Failure ]

X has Bernoulli Distribution

$$\boxed{P(X=x)} = p^x (1-p)^{1-x}$$

$$P(x) = \begin{cases} p & \text{if } x = 1 \\ q & x = 0 \end{cases}$$

② Multinomial Naive Bayes

- Discrete Count
- Multinomial E.g:- want count frequency of word in a particular document.

→ Multiple Probability distribution

Multinomial Distribution

$$P(x_1, x_2, \ldots x_k) = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}$$

| BG | O | A | B | AB |
|---|---|---|---|---|
| P | 0.44 | 0.42 | 0.1 | 0.04 |

6 indians

1 : O, 2 : A, 2 : B, 1 : AB

$$P(x_1 = 1, x_2 = 2, x_3 = 2, x_4 = 1) = \frac{6!}{1!\,2!\,2!\,1!} 0.4^1 \times 0.42^2 \times 0.1^2 \times 0.04^1$$

↓(k)

No. of occurance of particular outcome

PAGE NO.
DATE / /

Guassian Naive Bayes

PDF

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$\mu \rightarrow$ mean, $\sigma^2 \rightarrow$ variance, $\sigma \rightarrow$ std deviation

If data is discrete use Multinomial distribu
If -"- is Continuous use Guassian NaiveBayes
If features have continuous values

(1) Mean, Std deviation, variance
~~is a collections of facts, such as no. are, meal obse~~

Discrete Data [ Data takes Certain Value ]
eg:- No. of Students in a Class [Can't have half Student]
Rolling two diece [2, 3, 4, 5, 6, 7, 8, 9, 10, 11 & 12]

(2) Continuous Data
[Data Can take any Value [within a range]
A Person's ht. Could be any value [within range of human heights], not just certain fixed heights
· Time in a race : measure in fraction of second
· A dog's wt.
· The length of a leaf

Data: [Dog]

Qualitative
· He is a brown & black
has long hair
has lots of energy

Quantitative
Discrete
4 legs, 2 ears.
Continuous
weight 25.5 kg
565 mm tall