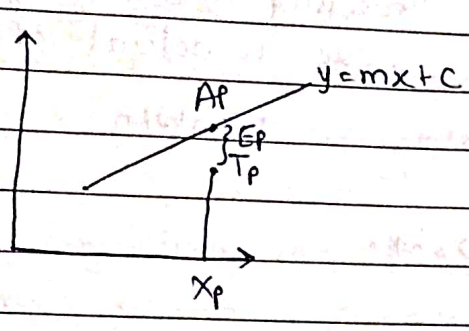
* Supervised learning algorithm
In this algorithm input & output is known.
Data is labelled (with teacher)

* Unsupervised learning algorithm
In this algorithm input & output are not known.
Data is unlabelled (without teacher)

Eg:



T_p = Target point
A_p = Actual point
E_p = Error at point P.

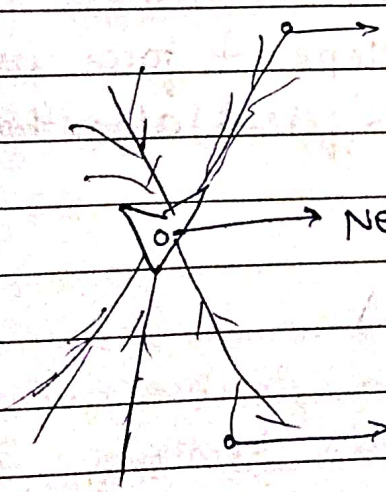
$$E_p = T_p - A_p$$

$$\text{MSE} = \sum_{i=0}^{p} \frac{(T_p - A_p)^2}{p}$$

Mean square error

· To minimize MSE we have to change m (slope) & c (intercept)
Here m & c are free parameters.

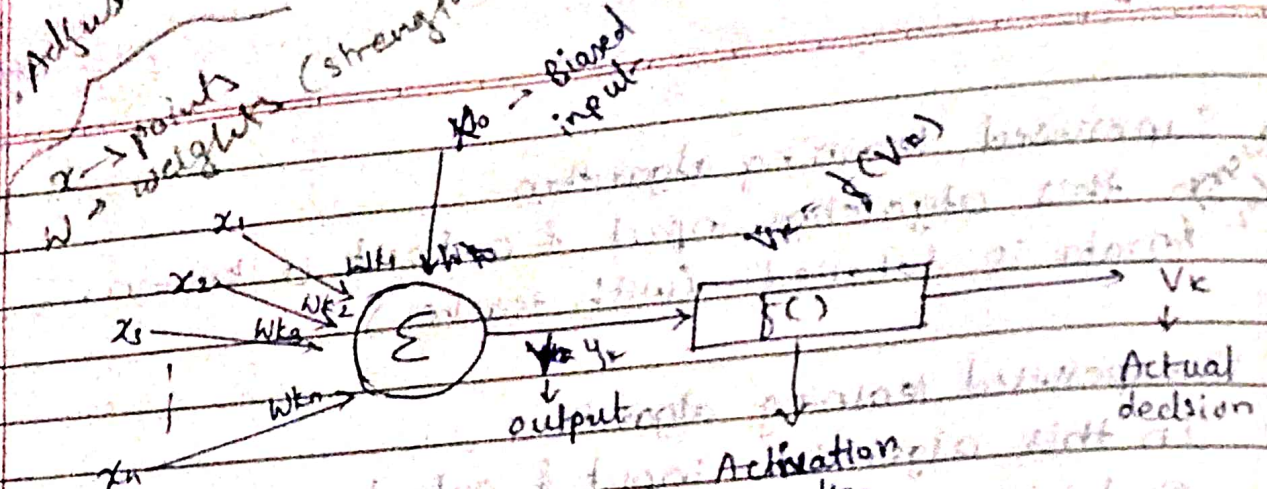
*



In out body we adjust strength of connection by free parameter

Adjust weights

x → points
w → weights (strength)



Processing to convert from processed value to actual value.

$$y = x_1 w_{k1} + x_2 w_{k2} + \dots + x_n w_{kn}$$

$$= \sum_{i=1}^{n} x_1 w_{k1} + x_0 w_{k0}$$

$$y_k = \sum_{i=0}^{n} x_1 w_{k1}$$

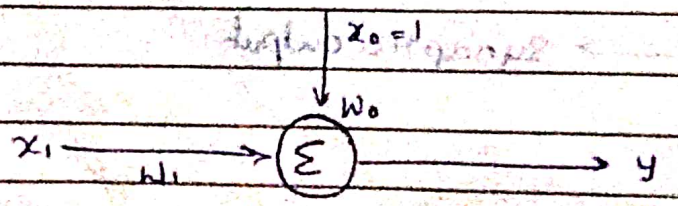
$$v_k = \begin{cases} 1 & y_k > \theta \\ 0 & \text{else} \end{cases}$$

- Gradient descent algorithm : Slope & force are in same direction.
- Steepest descent algorithm : Slope & force are in different algorithm

* $w_{new} = \eta\, m_{sE}\, x_i + w_{old}$

η → neta (learning rate)

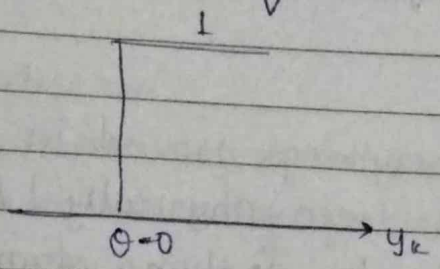
*



$$y = x_1 w_0 + x_0 w_0$$

$$y = x_1 w_1 + w_0 \quad \{x_0 = 1\}$$

* Types of activa" func"
1) Threshold activa" func"



Types of learning
1) Supervised learning: (inductive): Training data includes desired output.
Regression Classification.
linear non linear The algorithm learns relateonship bet" specific inputs & outputs based on training data & human feedback

2] Unsupervised learning: Training data doesnot include desired output.
clustering association
The algo. analyzes data for trends & patters without being given a specific output variable / human feedback.

3] Semi ~~surpvis~~ supervised: Training data includes few outputs.

4] Reinforcement:- ~~Re in~~ Rewards from sequence of action (eg. game)

- The problem of problem recognition is to identify the underlined structure within data.

- Intelligence is defined as ability ~~of~~ to comprehent to understand & profit ~~th~~ from experience

WORLD STAR
Date: 4
Page 27/7/22

- Plato said "Concept of abstract ideas are known to us a prior knowledge, through a mystic connec" with world.

- Pattern – Pattern is everything around in this digital world
  A pattern can either be seen physically / it can be observed mathematically by applying algorithms.
  Eg. Colors on clothes, speech pattern, etc
  In CS, patter is represented using vector feature values.

- Pattern recognition-
  PR is process of recognizing patterns by using machine learning algorithm.
  PR can be defined by classifica" of data based on knowledge already gained or on ~~stru~~ statistical information. extracted from patterns and / or their representation.

- Feature
  - Feature may be represented as continuous, discreate or discrete binary variables.
  - Numerical value given info. about object signal etc is called feature.
  - Broadly speaking feature are any extractable measure" used.
  - features may be :
    1) Symbolic (color)
    2) Numeric (Weight)
    3) Both.

- Feature when it is put into certain order is called feature recognition.

- Feature extraction.
  Feature extrac^n refers to process of transforming raw data into numerically feature that can be processed while preserving the info. in the original data set.

- It yeilds better results than applying machine learning directly into data

- Feature extrac^n can be accompolished manually or automatically.

- Feature selection imp features rakhna baki sab nikal dena
  Feature selec^n is the process of reducing no. of input variables when developing a predective model.

  It is desirable to reduce no. of input variables to both reduce the computational cost of modeling & in some cases, to improve performance of model.

  Feature selection is primarily focused on removing non-informative / redundant predictors from the model.

- Difference bet^n feature vector & feature space.
- A feature is a numerical / symbolic property of an aspect of an object.

- A feature vector is a vector containing multiple elements about an object.

- Putting feature vector for objects together can make up a feature space.

Scatter plot

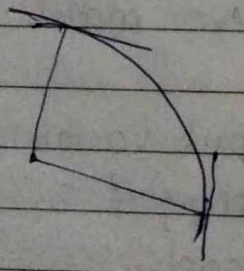
Scatter plot are graphs that represent the relationship bet^n 2 variables in a data set.

It represents data points on 2-D plane / cartesian system.

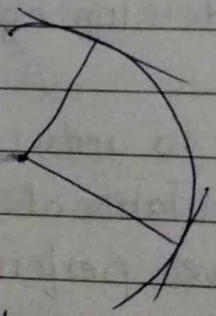
The independent variable or attribute is plotted on X-axis while dependent variable is plotted on y-axis.

Boundary based → Chain Coding Technique.
Region based.

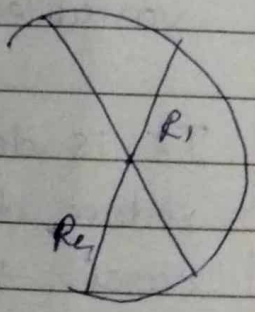
* Curve Pattern Recognition.



Radius 1
Center 1



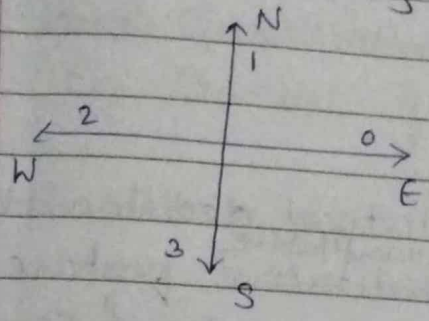
Radius 2
Center 2



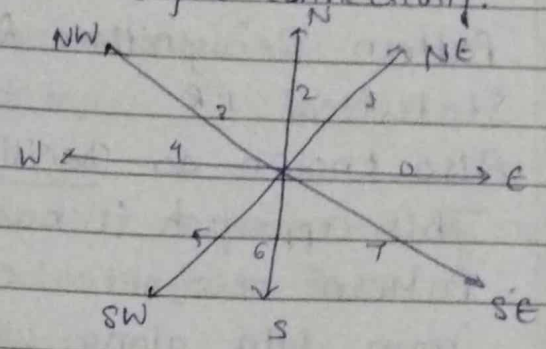
Chain Coding Technique.

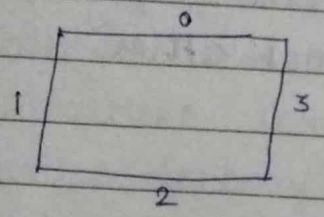
1) Four Connectivity
2) Eight Connectivity.

Four connectivity

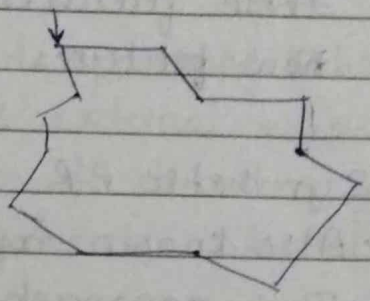


Eight Connectivity.





(0, 3, 2, 1)
Chain code

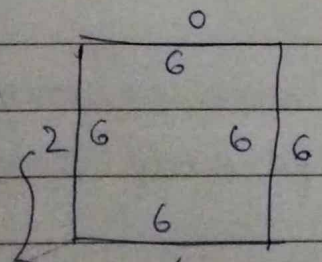


(0,7 0 6 7 5 3 4 3 1 2 1)
Chain-code

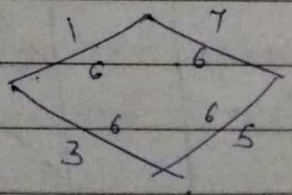
Translational invarient ∨ (Problem of origin rotaⁿ)
problem

Differential Chain Coding.

- To overcome translational invarient problem.
- If pattern is rotated then also it gives same difference (Advantage)



Betⁿ 2 & 4 there are 6 lines



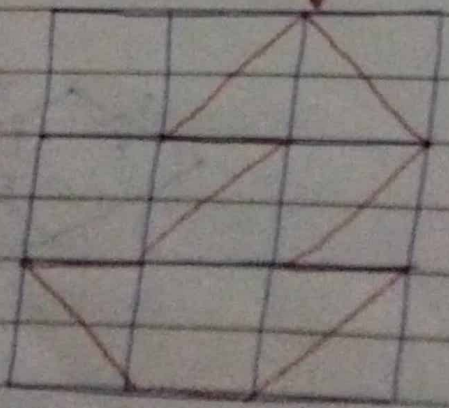
Pattern Recognition Approaches:

1) Statistical PR:
- Also known as Decision Theory.
- This approach is based on statistical decision theory.
- Pattern recognizes extracts quantitative features from data along with multiple samples & compares those features. However it doesn't touch upon how features are related to each other.

2) Synthetic PR
- Also known as Structural
- This approach is closest to how human perception works. It extracts morphological features from one data sample & checks how there are connected & related

3) Neural PR:
In this approach artificial neural network are utilized.
Compared to ones mentioned above, it allows more flexibility in learning & is the closest to naturally learning.

Q.



→ Chain code (8 connectivity)
7 5 0 5 4 3 0 1 4 1

Features.

1) ~~Qual~~ Quantitative data is numerical
Eg. Amount of rainfall / temp.

2) Qualitative data is discrete, with a finite no. of well-known values like weather descrip° in table

3) Morphological features.
This includes aspects of outward appearance (shape, structure, color, pattern, size) ie external morphology (or eidonomy) as well as form & structure of internal parts like bones & organs ie internal morphology (or anatomy).

Minimum perimeter length polygonization.

- Boundary is converted into set of "Concated Cells".
- Concatenated → Join / Combine.
- Cord try to shrink towards outward direction / inward direction until & unless reaches the limiting point.
- Moves only in one direction
- Moves towards resulting direction upto limiting points given by inward & outward direction.

Splitting of boundary

- Identify the set of points on arbitrary boundary along with this boundary can be split.
- The boundary points becomes ~~visit~~ vertices of the polygon.
- Criterion func^ : helps to decide part of point on arbitrary boundary along with which boundary can be splitted.

- Approximation 1 : Algorithm

S1: Initially decide 2 points along which we can divide the boundary.

S2: Draw straight line passing through these 2 points.

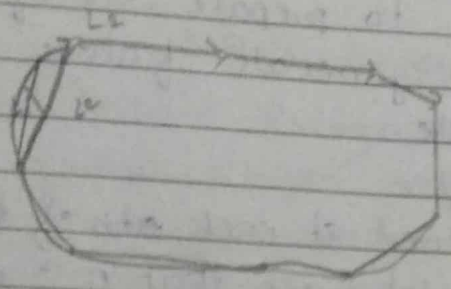
S3: Find max. distance on boundary point from the line in both segments.

S4: If max. distance is greater than threshold, then take maximum point as a next split point.

S5: Again find max. distance from the boundary & if distance is greater than threshold value then take as next split point else stop the algorithm.

Approximation 2: What is the area of portion of enclosed within straight line & original boundary.

Approximation 3:



Select an origin point.
Start drawing lines from that point ($L_1, L_2, L_3, \ldots$)
Suppose in the abv fig;
$L_2 > \theta$ (threshold val.)

then L_2 will be the final boundary.

Again repeat the same process of drawing lines from end point of L_2.

$$\theta_2 = \sum_{i=1}^{k} \alpha_i \theta_{i-1}$$

$$= \alpha_n \theta_{n-1} + \alpha_{n-1} \theta_{n-2} + \cdots + \alpha_k \theta_{n-k}$$

Feature vector.

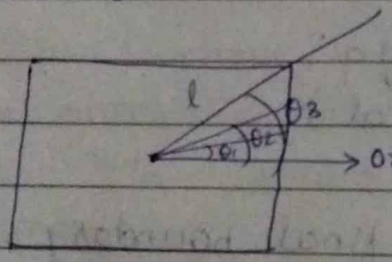
Polygon approxima^ of arbitrary boundary such polygon is not much helpful for PR purpose. Therefore, approximate with piecewise linear segments.

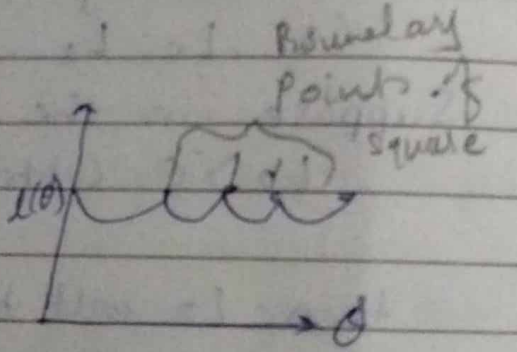
Finally, we have to prepare set of features vector. One approach to generate feature vector is

θ_n th angle.

- Inner angle sustained at each ~~oth~~ of the vertices.
- Autoregressive model, such that n^{th} angle can be represented by linear combina^ of k no. of previous angle.

Signature.





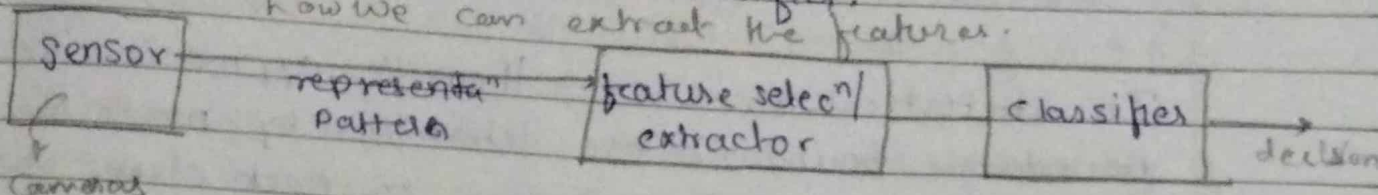
Main purpose of feature vector – Multiple features are kept in series

Statistical Pattern Recognition

WORLD STAR™ Date 3/8/22 Page 13

Block diagram of pattern classifier.

how we can extract the features.



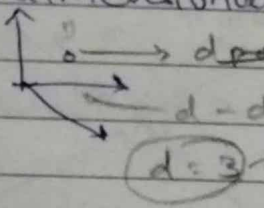
Pattern classifier..

Statistical Pattern Recognition.

Statistical pattern recogni" refers to use of statistics to learn from examples.

Eg of student in class 60 boys 20 girls probability of new student coming will be boy as per prior knowledge

- It means to collect observa", study & digest them in order to infer general rules or concept that can be applied to new unseen observation.
- In statistical approach, each pattern is represented in terms of d features or measurements & is viewed as point in d-dimensional space.



- The goal is to choose those features that allows patter vector belonging to different categories to occupy compact & disjoint region in d-dimensional feature space.
- The effectiveness of representa" space (feature set) is determined by how well patterns from different classes can be seperated.
- Given a set of training patterns from each class, objective is to establish decision boundaries in feature space

which seperates pattern belonging to ~~sam~~ different classes.

In statistical decision, theoritic approach, decision ~~boundaries~~ boundaries are determined by probablity distribu^n of pattern belonging to each class, which must either be either be specified or learned [41], [4

Gausian case & class dependence

Bayesian Network.

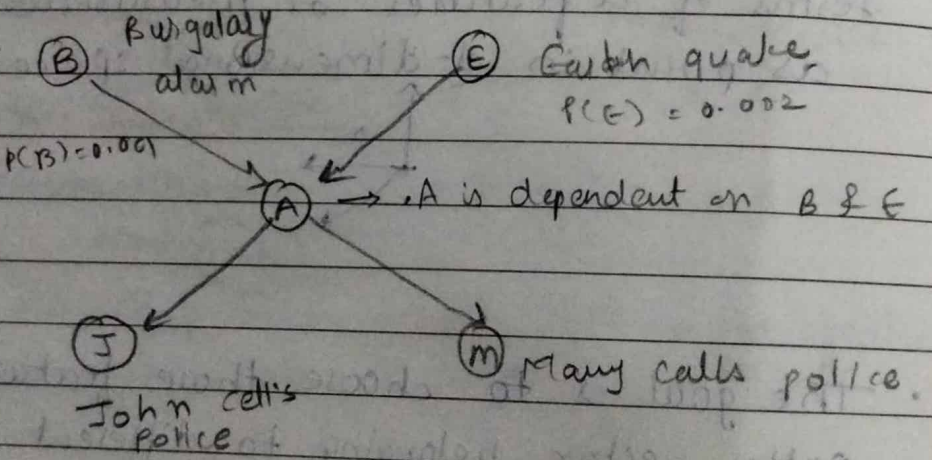
node → action

- Directed Acyclic Graph (DAG)
- Nodes: Set of random variables / Events
- Directed link: (X) → (Y)

X has direct influence on Y

- Conditional Probablity Table (CPT)

Eg:- B & E are independent events



For A

B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

Conditional probability

for J

A	P(J)
T	0.9
F	0.05

For M

A	P(m)
T	0.7
F	0.9

Here probablity depends on previous event hence sum of probablity is not one

Baye's rule:

Probablity of occuiance of event A & event B is given by probablity of A $P(A \wedge B)$ [probablity of A & B]

$$P(A \wedge B) = P(A|B) \cdot P(B)$$ If Already event B is occured & the probablity of occurance A is $P(A|B)$. This is known as conditional probablity.

$$P(A \wedge B) = P(B|A) \cdot P(A)$$ In this case probab. of B is dependent on probab. of A

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)}$$ (conditional probablity)

Joint probablity distribution.

$$P(x_1, x_2 x_3 -- x_n) = \prod_{i=1} P(x_i | \text{Prob. of parent})$$ (product)

In Above ex.

$$P(A, J, M) = P(A) * P(J|A) * P(M|A)$$
$$= P(A \wedge J \wedge M)$$

Q.1 $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$.?

$$= P(J|A) * P(M|A) * P(A|\neg B, \neg E) * P(\neg B) * P(\neg E)$$
$$= 0.9 * 0.7 * 0.001 * (1-0.001) * (1-0.002)$$
$$0.999 \qquad 0.998$$
$$= 0.000628$$

Q.2. $P(J)$.?

$$P(J) = P(J|A) * P(A) + P(J|\bar{A}) * P(\bar{A})$$
$$P(A) = P(A|B, E) * P(B \wedge E) +$$
$$P(A|\bar{B}\, E) * P(\bar{B} \wedge E) +$$
$$P(A|B\bar{E}) * P(B \wedge \bar{E}) +$$
$$P(A|\bar{B}\bar{E}) * P(\bar{B} \wedge \bar{E})$$

$= 0.9 * 0.001 * 0.002 +$
$0.29 * (1-0.001) * 0.002 +$
$0.95 * 0.001 * 0.998 +$
$0.001 * 0.999 * 0.998$

$= 0.0000018 + 0.000579 + 0.0009481$
0.00093
$= 0.002518$

$P(\bar{A}) = P(\bar{A} \mid B, E) * P(B \wedge E) +$
$P(\bar{A} \mid \bar{B}, E) * P(\bar{B} \wedge E) +$
$P(\bar{A} \mid B, \bar{E}) * P(B \wedge \bar{E}) +$
$P(\bar{A} \mid \bar{B}, \bar{E}) * P(\bar{B} \wedge \bar{E})$

$P(\bar{A}) = 1 - P(A)$

$= 0.05 * 0.001 * 0.002 +$
$0.71 * 0.999 * 0.002 +$
$0.05 * 0.001 * 0.998 +$
$0.999 * 0.999 * 0.998$

$= 0.0000001 + 0.001418 + 0.0000499 +$
~~0.00499~~ 0.9960049

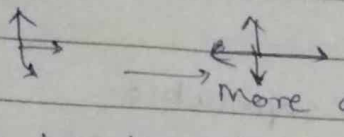
$= 0.9974729$

$P(J) = P(J \mid A) * P(A) + P(J \mid \bar{A}) * P(\bar{A})$
$= 0.9 * 0.002518 + 0.05 * 0.9974729$

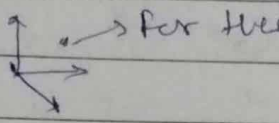
$= 0.022662 + 0.0498273$

$= 0.07253564$

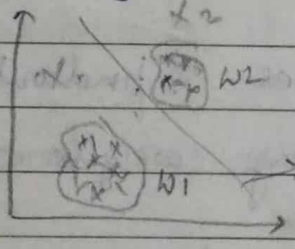
1) Dimentionality of feature vector dependent on problem under consideration.
2) Accuracy of descrip^n increases as we increase dimentionality of feature vector.



3) Increase in dimentionality increases complexity of problem
4) Dimension of PR problem based on the complexity of the problem.
   - Extract more no. of features
5) Once we represent pattern by pattern feature then this entire feature point is single point in dimensional feature space.



6) In PR problem we have to find out boundaries for patterns. Designing / training of classifier means to find out equation of decision boundary for given training samples. Therefore this is called supervised learning.



7) In above example, training samples are used to seperate using boundaries & it is possible to seperate using single line. Hence it is called as linearly seperable problem.

3-dim — Plane — 2 dim - line
multidimentional — Hyperplane

features → Input of classifier

Neural network always in bet^n 0 to 1

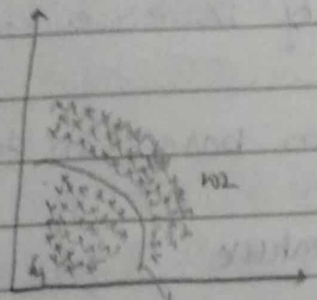
WORLD STAR™ Date Page 18

e) In case of 3-D problem classes can be sp seperated using plane. Hyperplane is required in multidimensional problem.

In all cases, eq^n of decision boundary is linear.

detailed data ? for given input o/p there

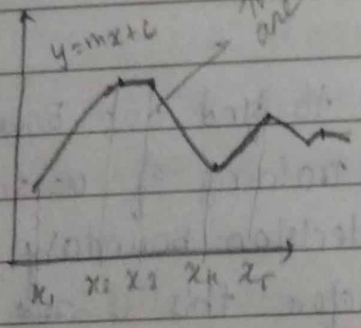
* If classes are not linearly seperable



Non linear decision boundary.

Quadratic classifier / cubic classifier

for more complex / complicated problem



In this input is only that is are elements

Linear approximation

$y = mx + c$

$y = W_1 x_1 + W_0$

$$O_j = f\left(\sum_{i=0} W_i x_i\right)$$

- Divide curve into lines & find out weights.
- we adjust these using weight matrix

weight matrix

W_0 W

* Joint probablity.

Q. The height of students are given below. Classify new students having height 3 feet.

P(B)	P(a)
6	6
7	5
5	3
7	2
6	3
5	3
4	4
3	5
5	2
6	1

$P(x|G) = P(a)$

Prior of probablity of boys $\Rightarrow 0.5$
Prior of probablity of girls $\Rightarrow 0.5$

Class conditional probablity $\Rightarrow$

$P(3|B) = P(3)/10 = 1/10 = 0.1$
$P(3|G) = P(3)/10 = 3/10 = 0.3$

Joint prob:-

$$P(G|3) = \frac{P(3|G) \cdot P(a)}{P(3)}$$

$$= \frac{0.3 \times 0.5}{0.2} = 0.75$$

$$P(3) = 0.1 \times 0.5 + 0.3 \times 0.5$$
$$= 0.2$$

$$P(B, 3) = \frac{P(3|B) \times P(B)}{P(3)}$$

$$= \frac{0.1 \times 0.5}{0.2} = 0.25$$

$$P(4, 3) > P(B, 3)$$
$$\therefore \text{It is a girl.}$$

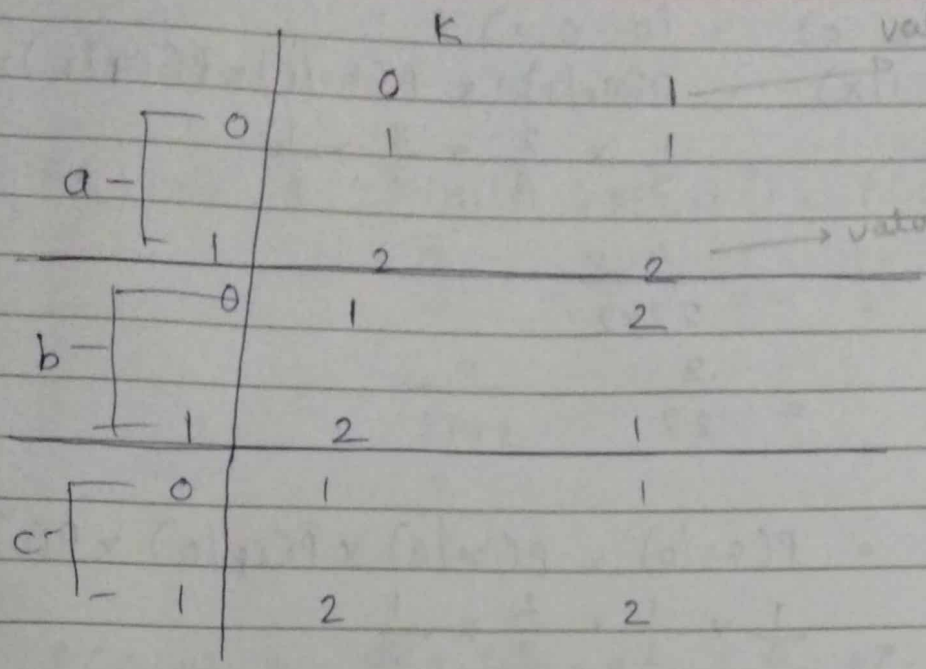
Q. Consider following dataset are features & t is the class dataset = a, b, c.

a	b	c	t
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1

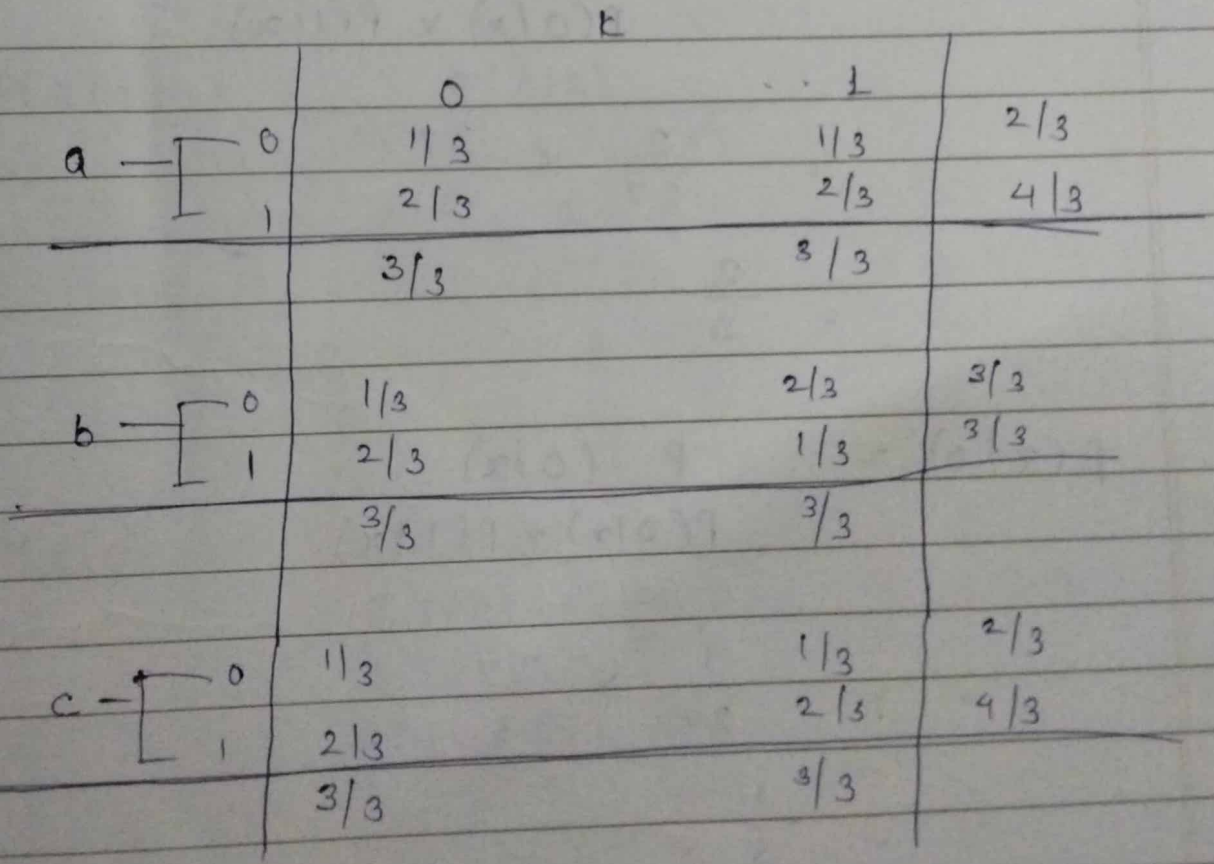
Classify the test instance given below into class 1 or 0 using Bayes classifier.

(a, b, c) = (0, 0, 1) (ab, c) = (1, 0, 0)

WORLD STAR™
Date :
Page 24



Since we have 3 I/P we divide each element by 3 & create likehood table.



$(a, b, c) = (0, 0, 1)$

$P(1|x) = P(a_0|1) \times P(b_0|1) \times P(c_1|1) \times P($

$= \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{2}$

Joint prob.

$= \frac{4}{27 \times 2}$

$= \frac{.2}{27}$

$P(0|x) = P(a_0|0) \times P(b_0|0) \times P(c_1|0) \times P(0)$

$= \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{2}$

Conditional

$= \frac{1}{27}$

$P(x|1) = \frac{P(1|x)}{P(0|x) \times P(1|x)}$

$= \frac{2}{27} \times \frac{27}{3}$

$= \frac{2}{3}$

$P(x|0) = \frac{P(0|x)}{P(0|x) + P(1|x)}$

$= \frac{1}{27} \times \frac{27}{3}$

$= \frac{1}{3}$

$P(x|1) > P(x|0)$

∴ It belongs to class 1

$(a, b, c) = (1, 0, 0)$

$$P(1|x) = P(a,|1) \times P(b_0|1) \times P(c_0|1) \times P(1)$$

$$= \frac{2}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{2}$$

$$= \frac{4}{27 \times 2}$$

$$= \frac{2}{27}$$

$$P(0|x) = P(a,|0) \times P(b|0) \times P(c|0) \times P(0)$$

$$= \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2}$$

$$= \frac{1}{27}.$$

$$P(x|1) = \frac{P(1|x)}{P(1|x) + P(0|x)}$$

$$= \frac{2}{27} \times \frac{27}{3}$$

$$= \frac{2}{3}$$

$$P(x|0) = \frac{P(0|x)}{P(0|x) + P(1|x)}$$

$$= \frac{1}{27} \times \frac{27}{3}$$

$$= \frac{1}{3}$$

$P(x|1) > P(x|0)$

So it belong to class 1

$(a, b, c) = (1, 0, 1)$

$P(1|x) = P(a_1|1) \times P(b_0|1) \times P(c_1|1) \times P(1)$

$= \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{2}$

$= \frac{4}{27}$

$P(0|x) = P(a_0|0) \times P(b_0|0) \times P(c_1|0) \times P(0)$

$= \frac{2}{3} \times \frac{1}{3} \times \frac{7}{3} \times \frac{1}{2}$

$= \frac{2}{27}$

$P(x|1) = \frac{P(1|x)}{P(1|x) + P(0|x)}$

$= \frac{4}{27} \times \frac{27}{6} = \frac{4}{6} = \frac{2}{3}$

$P(x|0) = \frac{P(0|x)}{P(0|x) + P(1|x)}$

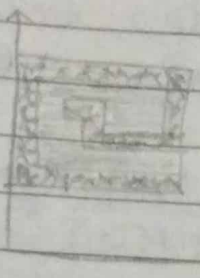
$= \frac{2}{27} \times \frac{27}{6} = \frac{1}{3}$

$P(x|1) > P(x|0)$

∴ It belongs to class 1.

WORLD STAR™
Date: 17/8/22
Page: 27

Bayes decision Theory. (Boxes)



Linear $y = mx+c$.
Non linear quadratic.

Let. ω_1 → Probablity of acceptance $P(\omega_1)$
ω_2 → Probablity of rejection $P(\omega_2)$

Rule 1 :-
If $P(\omega_1) > P(\omega_2)$ then class (ω_1) (x = New input)

Rule 2 :-
If $P(\omega_2) > P(\omega_1)$ then class (ω_2)

A priori probablity (APP)
↳ It is completely based on History

APP + some features (x).

x → Can be have various values.

Class conditional PDF

$P(x|\omega_1)$ → Probablity density func^n (PDF) of variable x for the object that belongs to ω_1

$P(x|\omega_2)$ → Probablity density func^n (PDF) of variable x for the object that belongs to ω_2.

→ NO } classes (error-free rejection)

Joint probability

$P(\omega_i, x) = P(\omega_i/x)\, P(x)$

$= P(x/\omega_i)\, P(\omega_i)$ → By probability theory

$P(\omega_i/x) \cdot P(x) = P(x/\omega_i)\, P(\omega_i)$

$P(\omega_i/x) = \frac{P(x/\omega_i)\, P(\omega_i)}{P(x)}$

$P(x) = \sum_{i=1}^{2} P(x/\omega_i)\, P(\omega_i)$

Probability input ↑

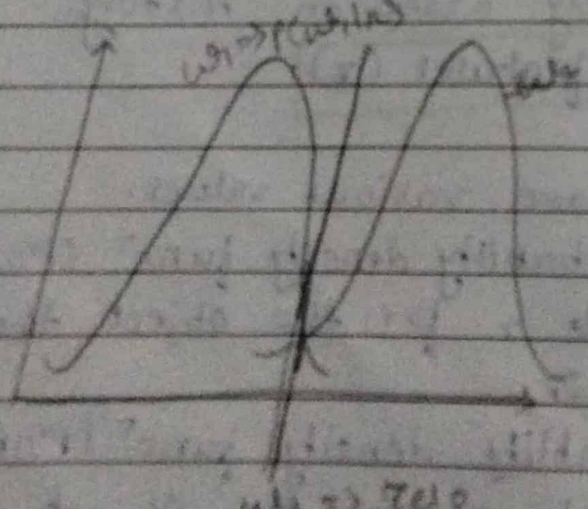
$\begin{cases} P(\omega_1/x) > P(\omega_2/x) \Rightarrow x \text{ Belongs to } \omega_1 \\ P(\omega_2/x) < P(\omega_2/x) \Rightarrow x \text{ Belongs to } \omega_2 \end{cases}$

$\therefore P(x/\omega_1) \cdot P(\omega_1) > P(x/\omega_2) \cdot P(\omega_2)$ → ω_1

(Major probability)

$P(x/\omega_1) \cdot P(\omega_1) < P(x/\omega_2) \cdot P(\omega_2)$ → ω_2

- If $P(\omega_1)$ & $P(\omega_2)$ is same, the decision depends on class conditional probability density fn.



- $P(\omega_2|x)$ should be zero (left side)
If it has some value, then this finite probability error occurs. & vice versa.

Generalizeⁿ of Baye's theory.

1) Use more than 2 states of nature (classes)
2) In this classificaⁿ, use more than 1 feature.
If there are more than 1 feature then it is feature vector.
3) Allow other actions other than merely deciding steps of nature.
4) Introduce a loss function which is more general than probablity of error.

Let us consider,

C no. of classes $\{\omega_1, \omega_2, \omega_3 \dots \omega_c\}$

a - No. of actions $\{\alpha_1, \alpha_2, \alpha_3, \dots \alpha_a\}$

$\lambda(\alpha_i|\omega_j)$ → Loss function.
} Loss incurred for taking action α_i when true state nature is ω_j {

- $x \Rightarrow$ d dimensional feature vector.

for x feature vector, we considers acⁿ α_i

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) \cdot P(\omega_j|x)$$

↳ Conditional Risk / Risk / Accepted loss.

classifying Risk x by taking acⁿ α_i

Minimum risk classifier

i → Act^n
j → class α_i → Action

WORLD STAT

- Minimum risk classifier.

ω_1 ω_2 → Class

α_1 α_2 → Actions.

$\lambda(\alpha_i | \omega_j) \Rightarrow \lambda_{ij}$

$R(\alpha_1|x) = \lambda_{11} P(\omega_1|x) + \lambda_{12} P(\omega_2|x)$

$R(\alpha_2|x) = \lambda_{21} P(\omega_1|x) + \lambda_{22} P(\omega_2|x)$

$\omega_i \Rightarrow$

$R(\alpha_1|x) < R(\alpha_2|x)$

$[\lambda_{11} P(\omega_1|x) + \lambda_{12} P(\omega_2|x)] < [\lambda_{21} P(\omega_1|x) + \lambda_{22} P(\omega_2|x)]$

$(\lambda_{21} - \lambda_{11}) P(\omega_1|x) > (\lambda_{12} - \lambda_{22}) P(\omega_2|x)$

$\lambda_{21} P(\omega_1|x) > \lambda_{12} P(\omega_2|x)$.

Minimum Error Rate Classifier.

We have to maximize the probability to decrease the risk.

$\alpha_i \Rightarrow$ True State of nature is ω_i.

$\lambda(\alpha_i|\omega_j)$ Loss func^n $= \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$ If $i \neq j$ then true condition (High loss)

$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) \cdot P(\omega_j|x)$

$= \sum_{i \neq j} \lambda(\alpha_i|\omega_j) \cdot P(\omega_j|x)$ (Not possible for one value for no possible condition)

- $\sum_{i \neq j} P(\omega_j|x)$ (Not possible condition)

$$R(\alpha_i|x) = 1 - P(\omega_i|x)$$

This will minimize the risk.

Discriminant function.

- This classifier will compute the number equals to classes. ie if no. of classes are c then we have to calculate c no. of ~~risk~~ accepted risk / loss function / risk function.

are called as discriminant function

$$\begin{cases} g_1(x) = P(\omega_1|x) \\ g_2(x) = P(\omega_2|x) \\ \vdots \\ g_c(x) = P(\omega_c|x) \end{cases}$$

- Whichever discriminant func^n gives max. val. we will put x into that class.

$\Rightarrow g_i(x) > g_j(x)$ This condition is true only except
$\Rightarrow \forall j \neq i$ All values of i should not be equal to j.

- If ~~this~~ this condition is true then given feature vector belong to class ω_i.

- $\sum_{i \neq j} P(w_j|x)$ (Not possible condition)

$$R(x_i|x) = 1 - P(w_i|x)$$

This will minimize the risk.

Discriminant function. (working func^n of every class)

- This classifier will compute the number equals to classes. ie if no. of classes are c then we have to calculate c no. of ~~risk~~ accepted risk / loss function / risk function.

- are called as discriminant function

$$\begin{cases} g_1(x) = P(w_1|x) \\ g_2(x) = P(w_2|x) \\ \vdots \\ g_c(x) = P(w_c|x) \end{cases}$$

- Whichever discriminant func^n gives max. val. we will put x into that class.

- $g_i(x) > g_j(x)$ This condition is true only except
  - $\forall j \neq i$ All values of i should not be equal to j.
  - If this condition is true then given feature vector belong to class w_i.

Minimum Error Classifier

$g_i(x) = P(w_i|x)$

For minimum error classifier, here require to optimize risk func^n to minimum & it is possible by making $P(w_i|x)$

As a result,

($R(\alpha_i|x) = 1 - P(w_i|x)$), risk factor will be reduced.

$f(g_i(x)) = 2.g_i(x)$

Monotonically increasing function.

$g_i(x) = P(w_i|x)$

$g_i(x) = \ln P(x|w_i) + \ln P(w_i)$

Q. The height of students are given below. Classify new students having height 3 feet

Boys	Girls
6	6
7	5
5	8
7	2
6	3
5	3
4	4
3	5
5	1
6	2

Prior probability { boys => 0.5
Prior probability { girls => 0.5

Class conditional probability =>
$P(3|B) = P(3)/10 = 1/10 = 0.1$
$P(3|G) = P(3)/10 = 3/10 = 0.3$

Joint prob:-

$$P(G|3) = \frac{P(3|G) \cdot P(G)}{P(3)}$$

$$= \frac{0.3 \times 0.5}{0.2} = 0.75$$

$$P(3) = P(3|B) \times P(B) + P(3|G) \times P(G)$$
$$= 0.1 \times 0.5 + 0.3 \times 0.5$$
$$= 0.2$$

$$P(B,B) = \frac{P(3|B) \cdot P(B)}{P(3)}$$

$$= \frac{0.1 \times 0.5}{0.2}$$

$$= 0.25$$

$\therefore P(G,3) > P(B,3)$.

$\therefore$ It is a girl.

Gaussian Parametric Estimation of Prob. density function.

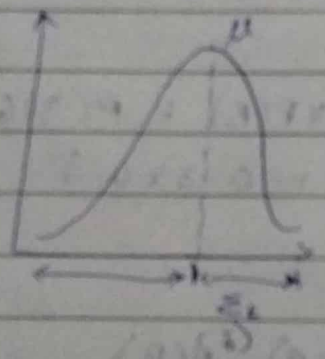
↳ MLE (Maximum Likelihood Estimation)

Gaussian Distribution:

$\omega_1 = \omega_1 \dots \omega_c$

$P(x|\omega_j) \sim N(\mu_j, \Sigma_j)$. $P(x|\omega_j)$ ← Gaussian distⁿ

Avg ; Standard deviation (δ) ; i/p feature vector belongs to ω



$$P(x|\omega_i) = \frac{1}{\delta_j \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_j}{\delta_j}\right)^2}$$

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} x_k.$$

$$\Sigma_x(\delta) = \frac{1}{n} \sum_{k=1}^{n} (x_k - \hat{\mu}) \times (x_k - \hat{\mu})^T$$

Reinforcement - At every stage will get some reward.

WORLD STAR
Date: 38
Page: 7/9/22

Learning Rules.

1] Supervised - Labeled data / with teacher
2] Unsupervised - Unlabelled data / without teacher.
3] Reinforcement - Rewards.

1) Delta learning ~~Rule~~ Rule.
2) Hebbian learning Rule
3) Memory based learning Rule
4) Competative learning Rule
5) Outstar learning Rule
6) Boltzman learning rule.

① Delta learning rule:

- Widraw Hoff
- Error correc^n LR
- ~~Erro~~ Mean square learning.
- Adaptive learning Elements (Adaline)

Let $x_1, x_2 \dots x_n$ are inputs connected with the strength of connections also called as weights $w_1, w_2 \dots w_{kn}$

w_{kj} is weight of connection from j^{th} input to k^{th} processing unit.

$$v_k = \sum_{i=1}^{n} x_i w_{ki} + x_0 w_{k0}$$

for p^{th} input

$$E_{k(p)} = T_{k(p)} - Y_{k(p)}$$

$E_{k(p)}$: Error at point p
$T_{k(p)}$: Targeted point (Swagat)
$Y_{k(p)}$: Actual point (Sarat bauji)

- $E(n) = \frac{1}{n} \sum_{p=1}^{n} (T_{k(p)} - Y_{k(p)})^2$

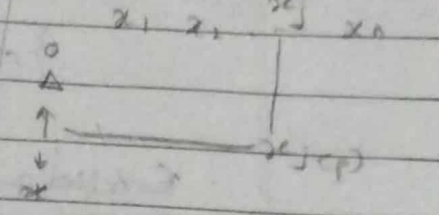
Total error

- $\Delta W_{kj(p)} = \eta E_{k(p)} \cdot x_{j(p)}$

Change in weight, learning rule, Error, x_j inputs move ($x_1, x_2 \ldots x_p \ldots x_j$), p^{th} pattern

- $W_{kj} = W_{kj(o)} + \Delta W_{kj(p)}$

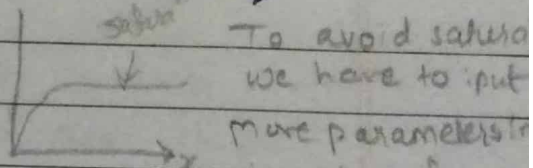
old W_{kj}



② Hebbian Learning Rule. / Activity Product Rule.

If one neuron simulates second neuron & frequently contributes in its activation, the synaptic association bet^n these 2 neurons is strengthened & second neuron become extra sensitive to simuli from 1st neuron.

sum func^n of product



$$\Delta W_{kj(p)} = f(x_{j(p)} \cdot y_{k(p)})$$

$$\Delta W_{kj(p)} = \eta (x_{j(p)} \cdot y_{k(p)})$$

ϕ = forgetting factor

$$\Delta W_{kj(p)} = \eta (x_{j(p)} \cdot y_{k(p)}) - \phi \eta (x_{j(p)} \cdot y_{k(p)})$$

0-1 range

[0.1 – 0.001]

③ Memory based learning / Instance based learning

Unsupervised learning

$\overline{X} = \{X_1, X_2, X_3 \ldots X_m\}$ m dimensional

elements

m ⇒ dimension

d = training patterns. / no. of feature vectors.

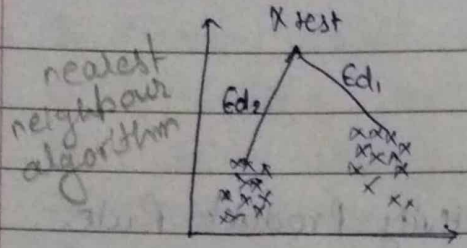
nearest neighbours } k-nearest neighbour
2 algo. used in ML for

$\{x\}_{i=1}^{N}$

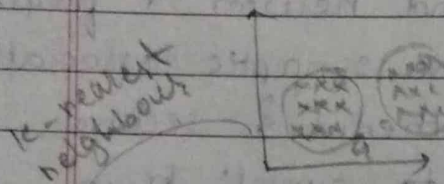
$\{d_i\}_{i=1}^{N}$

- ~~Euclidia~~ $\{\overline{x_i}, d_i\}_{i=1}^{N}$

- Euclidia distance.



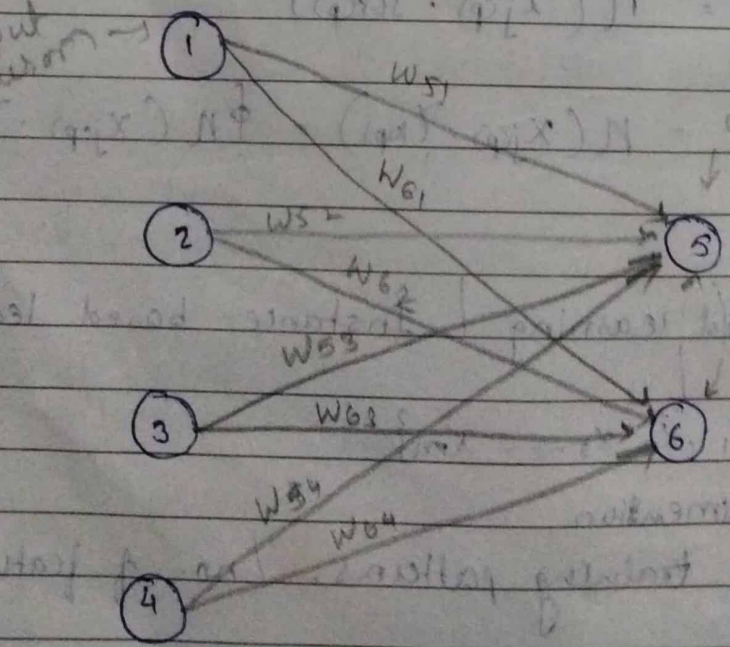
$Ed_1 < Ed_2$ → x test will classify into Ed_2.
x test.



If there is boy with long hairs, then that boy will be outlier

→ To minimise the error, we have to find out mean of the inputs. then this is known as k-nearest neighbour algo.

(4) Compititive learning Rule.



- inhibitory in nature (winners will not support each other)

WORLD STAR™
Date :
Page :

- Condition:-
  $w_{S1} + w_{G1} = 1$

  ie $\sum_{j=1}^{n} w_{kj} = 1$

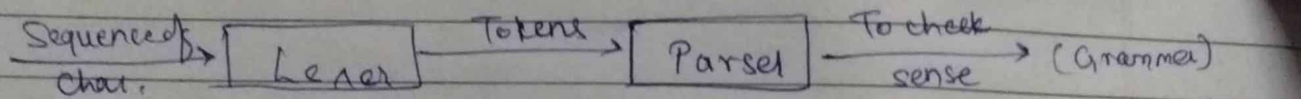
- Winner will be favour.

- In this algo. winner takes all reasons.

Parsing

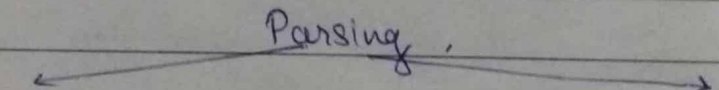
- Parsing is a process of driving string from the given grammar.
- Grammar. Study of the way words are used to make sentences.
- Rule of language governing the sounds, words, sentences & other elements as well as their combination of interpretation (Grammar).

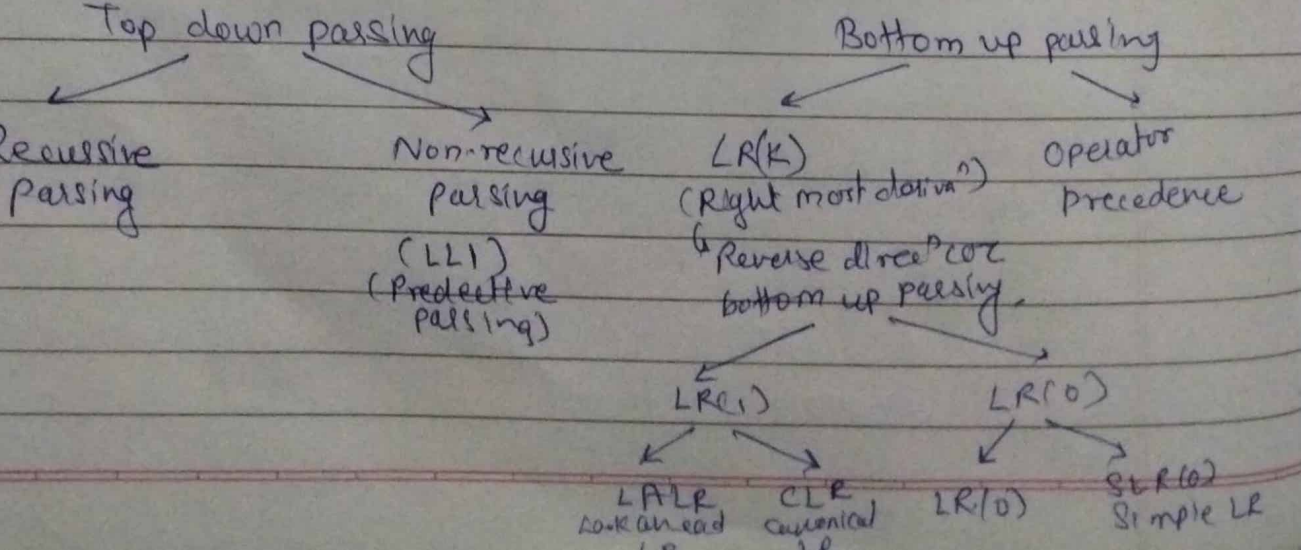
- Lexical Analysis.
- It is a ~~person~~ of ~~can~~ process of converting sequence of characters into sequence of steps.
- In computer science, lexical Analysis is process of converting a sequence of characters. (Such as in computer page / web page). into a sequence of lexical tokens (string with page).
- Lexer, Tokanizer or scanner)



CFG (Context free Grammer)

Parse tree / Syntax tree / ~~drive~~ tree
derive

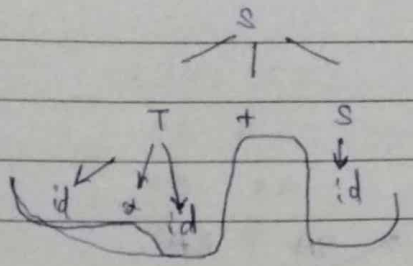




* Ambiguous grammer: A CFG is said to ~~abb~~ ambiguous if there exists more than one derivation tree for the given input string ie more than one LMDT / more than one RMDT.

* $S \rightarrow T + S \mid T$

  $T \rightarrow id * T \mid id \mid S$



id = id * id + id.

* Unambiguous grammer:
  CFG (Context Free Grammer)

CYK (Passing) Algorithm.

↓ Cocke Younger Kasmi

$S \rightarrow AB$

$A \rightarrow BB \mid a$

$B \rightarrow AB \mid b$

- CFG —form→ CNF / GNF
  - CNF: Chomsky Normal form
  - GNF: Greibach Normal form.

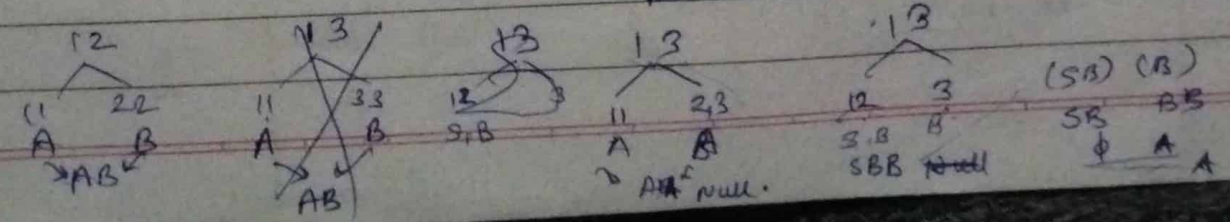
- CYK only support CNF.
- It is universal (~~all~~ applicable for all CNF)

Eg: Pattern → abbb.

a	b	b	b
1	2	3	4

1 to 1: (1,1) a → A; (2,2) b → B

	4	3	2	1
1	S,B	~~S~~ A	S,B	A
2	S,B	A	B	
3	A	B		
4	B			



WORLD STAR
Date :
Page :

2 3 4

22 | 4

23 44

A B

AB

S,B

2 34

22 34

B A

BA

Null. (ϕ)

2 34

12 34

S,B A

(SB) (A) → SA BA

ϕ ϕ

1 2 3 4

(11) 234

A S,B

(AS) (AB)

ϕ S,B

(125) 44

A B → S,B

a b a ba

a b a b a

1 2 3 4 5

	5	4	3	2	1
1	Null	S,B	A	S,B	A
2	Null	AB	Null	B	
3	A	S,B	A		
4	Null	B			
5	A				

S → AB

A → BB/a

B → AB/b

4,5

44 55

B A

23 44

BA BB

ϕ A

1,2,3 12 3

12 33 11 23

S,B A A ϕ

SA BA A

ϕ

ϕ

2,3,4

23 44 22 34

ϕ B B S,B

B (BS) (BB)

A

3 4 5

33 45 34 55

A ϕ S,B A

A SA BA

ϕ ϕ

ϕ

WORLD STAR™
Date :
Page :

1 2 3 4

(11) (234)

A AB

AB AB

ϕ S,B

(123) (44)

A B

AB

S,B

2 3 4 5

(22) (3,45)

B A

ϕ

(234) (55)

AB A

AA BA

ϕ ϕ

1 2 3 4 5

(11) (2345)

A ϕ

A

(1234) (55)

S,B A

SA BA ϕ

WORLD STAR™
Date 3/11/22
Page

Syntactic Analysis System

1) Phrase structure grammar
2) The transformational grammar
3) Transition network grammar.

1) Phrase structure grammar.

Purely syntactic analysis of language does not do very much in language processing.

- Syntax adjust to a more complete analysis process. Syntax is better understood ~~that~~ than most other linguistic procedure.
- Linguistics are noted for their work on generative grammars.
- Generative grammars are used to generate gramatically correct statements.
- Generative grammars have many property of simplicity.
- Particular type of grammar is known as ~~ph~~ phrase structure
- Simple & usable both for sentence generation & for sentence analysis.

Consider the rewrite rules of the form

S → A + B

↳ Statement

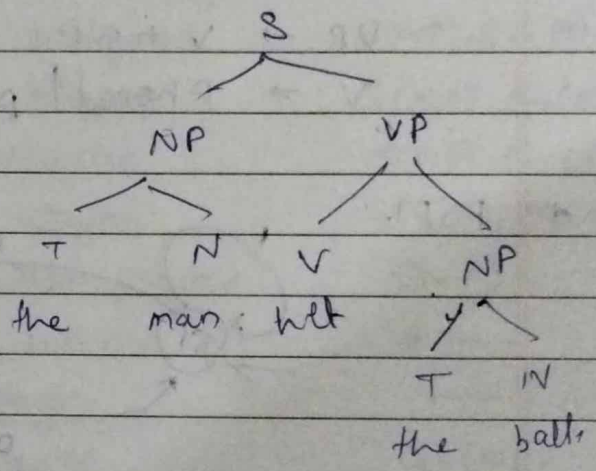
- Variable S can be rewritten as A followed by B (S followed by A & B)
- The symbol plus seperates the variables.
- When presenting rewrite rules capital letters denote non terminal elements. ie elements that can be rewritten further, by appearing on the left side of some rewrite rule.

Where the bar stands for logical connective: OR.
The result would be huge grammar which wouldnot be simple.

Eg: the man hit the ball.
the man hit the man.
the ball hit the ball.
the ball hit the ball.
the man took the man.
the man took the ball.
the ball took the man.
the ball took the ball.

Rules:
* S → NP + VP
* NP → T + N.
  T → the
  N → man, ball, etc. →
* VP → V + NP
  V → hit / took.

Derivation tree.



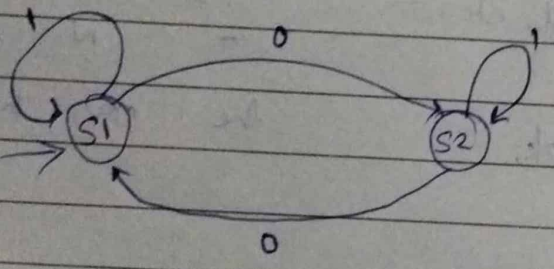
Context free
Grammar such that of expression 1 is known as context free phrase structure grammar bcoz non terminal symbols of LHS of the rewrite rules can be replaced by the right sides of the rules regardless of the context in which these symbols may appear. ie no restrictions apply to any rewrite rule.

John phoned Mary.
Mary phoned John.
John phoned John.
Mary phoned Mary.

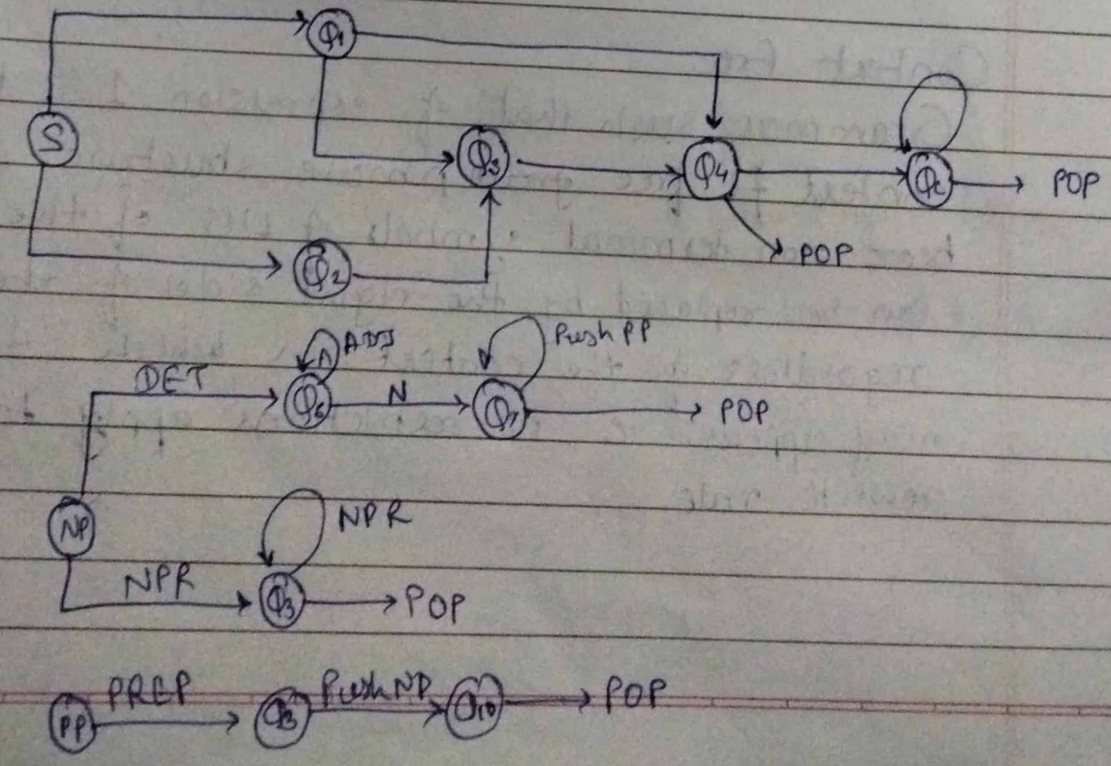
John phoned up Mary
John phoned up John.
Mary phoned up John.
Mary phoned up Mary

S → NP + VP.
NP → John | Mary.
VP → V + NP
V → phoned | phoned up.

* FSM.



*



The tall man in the Stetson is John Wayne.

	Current State	I/P string rec.	Stack.
1	S	The tall man --	--
2.	NP	The tall man --	S(NP)
3	Q6	tall man in ---	S(NP)
4	Q6	man in the ---	S(NP)
5	Q7	in the Stetson --	S(NP)
6	PP	in the Stetson is ---	S(NP), Q7(PP)
7	Q9	the Stetson is John --	S(NP), Q7(PP)
8	NP	the Stetson is John --	S(NP), Q7(PP), Q9(NP)
9	Q6	Stetson is John Wayne	S(NP), Q7(PP), Q9(NP)
10	Q8	is John Wayne	S(NP), Q7(PP), Q9(NP)
11	POP	is John Wayne	S(NP), Q7(PP), Q9(NP)
12	Q10	is John Wayne	S(NP), Q7(PP)
13)	POP	is John Wayne	S(NP), Q7(PP)
14	POP	is ~~John~~ John Wayne	S(NP)
15	Q1	is John Wayne	S —
16	Q4	John Wayne	←
17	NP	John Wayne	Q4(NP)
18	Q8	Wayne	Q4(NP)
19	Q8	—	Q4(NP)
20	POP	—	—
21	Q4	—	—
22	Pop	Accept	—