Unit-II

Support Vector & kernel Methods.

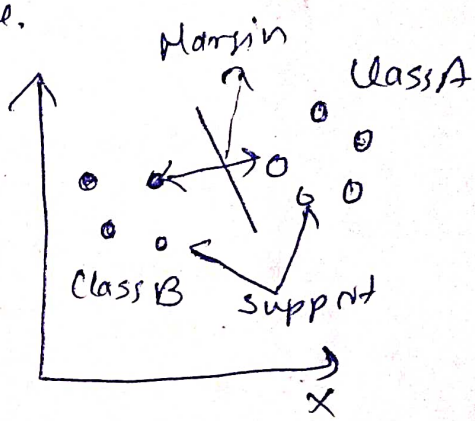
SVM :- flexible supervised Machine learning algorithm which are used both for classification & regression.

→ generally used in classification problem.

Support Vector Machine model is basically a representation of different classes in a hyperplane in multidimensional space.

→ The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized.

→ The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane.



SVM Kernals:-

SVM algorithm is implemented with kernels that transforms an input data space into the required form.

→ SVM uses a technique called the kernal trick in which kernel takes a low dimension input space and transform it into a higher dimensional space.

→ The following are some of the types of kernel methods used by SVM.

- **Linear kernel**
  It can be used as a dot product between any two observation
  formula :- $k(x, x_i) = \text{sum}(x * x_i)$

- **Polynomial kernel**
  It is more generalized form of linear kernal & distinguish curved or nonlinear input space.
  formula :- $k(X, X_i) = 1 + \text{Sum}(X * X_i) \wedge d$
  Here d is the degree of polynomial, which we need to specify manually in learning algorithm.

- **Radial Basis function (RBF) kernel**
  RBF kernel, mostly used in SVM classification, maps input space in indefinite dimension space.
  formula :- $K(X, X_i) = \exp(-\text{gamma} * \text{Sum}(X - X_i \wedge 2))$
  gamma ranges from 0 to 1, default value is 0.1

Ex:- Vector quantization & Kohonen maps (self organizing maps) --

Principles of CL:-

Three basic elements to a Competitive learning rule:

- A set of neurons that are all same, respond to a given set of input.
- A Limit of each neuron
- A mechanism, that permit the neurons to compete for the right.

→ hidden layer → Competitive layer

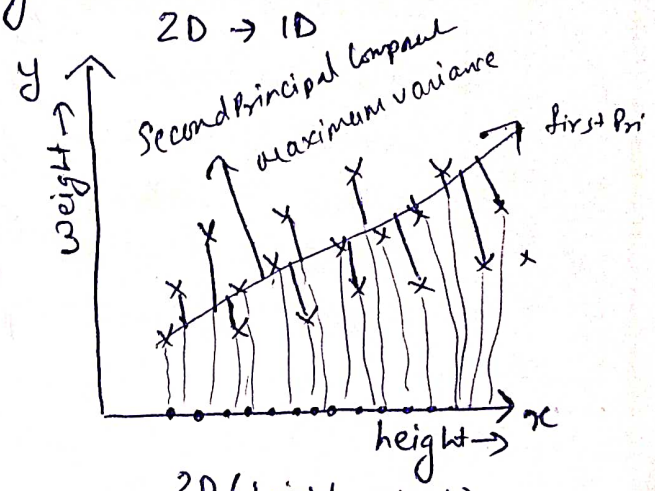
Principal Component Analysis :- "Dimensional Reduction"

higher dimension → lower dimension

2D → 1D

→ The line which has maximum number of data points known as First Principal Component

→ The line that is perpendicular on FPC is known as SecondPC



2D (height, weight)

↓

1D (First Principal Component)

Procedure for performing PCA

Step 1: Data Set $n \to < n$

Feature	Ex 1	Ex 2	...	Ex N
x_1	x_{11}	x_{12}	...	x_{1N}
x_2	x_{21}	x_{22}	...	x_{2N}
⋮	⋮	⋮	...	⋮
x_n	x_{n1}	x_{n2}	...	x_{nN}

Step 2: Compute the mean of the Variables

Mean of X_i

$$X_i = \frac{1}{N}(x_{i1} + x_{i2} + \dots x_{iN})$$

Step 3: Calculate the Covariance Matrix

Covariance of all the ordered pair (X_i, X_j)

$$\rightarrow Cov(X_i X_j) = \frac{1}{N-1} \sum_{K=1}^{N} (X_{iK} - \bar{X}_i)(X_{jK} - \bar{X}_j)$$

→ Construct $n \times n$ Matrix S called

$$S = \begin{bmatrix} Cov(X_1, X_1) & Cov(X_1 X_2) & \cdots & Cov(X_1 X_n) \\ Cov(X_2 X_1) & Cov(X_2 X_2) & \cdots & Cov(X_2 X_N) \\ \vdots & \vdots & \cdots & \vdots \\ Cov(X_n X_1) & Cov(X_n X_2) & \cdots & Cov(X_n X_N) \end{bmatrix}$$

Step 4: Calculate the eigen values and normalize eigen vectors of the covariance matrix

→ To find eigen value, solve equation

$$\det(S - \lambda I) = 0$$

S = Covariance matrix

λ = root of this equation

I = Identity matrix

We get n roots $\lambda_1, \lambda_2 \ldots \lambda_n$.

Which are eigen values such that $\lambda_1 > \lambda_2 > \ldots > \lambda_n$

→ for each eigen values the corresponding eigen value is a vector

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad n \times 1$$

Such that

$$(S - \lambda' I) U = 0$$

λ' = eigen value

→ Normalize the eigen vector

→ Divide the vector, U by its length.

i.e. Normalize eigen vector

$$e_i = \frac{U_i}{\|U_i\|}$$

where $\|U\| = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}$

Step 5:- Derive new data set

Q ~~PCA~~ data = { 2, 3, 4, 5, 6 7; 1, 5, 3, 6, 7, 8 }

Compute principal component using PCA

Q.2 Step 1 : Data Set

Feature	Ex 1	Ex 2	Ex 3	Ex 4
x	4	8	13	7
y	11	4	5	14

No. of features, $n = 2$

No. of samples, $N = 4$

Step 2: Mean of Variable

$$\bar{x} = \frac{4+8+13+7}{4} = 8$$

$$\bar{y} = \frac{11+4+5+14}{4} = 8.5$$

Step 3: Computation of Covariance Matrix

Ordered pairs

(x, x), (x, y), (y, x) (y, y) (x, y).

1) Covariance of all ordered pairs $\{(x_i - \bar{x})^2\}$

$$Cov(x, x) = \frac{1}{N-1} \sum_{k=1}^{N} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$$

$$= \frac{1}{4-1}\left[(4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2\right]$$

$$= 14$$

$$Cov(x, y) = \frac{1}{4-1}\left[(4-8)(11-8.5) + (8-8)(4-8.5) + (13-8)(5-8.5) + (7-8)(14-8.5)\right]$$

$$= -11$$

$$Cov(y, x) = Cov(x, y) = -11$$

$$Cov(y, y) = \frac{1}{4-1}\left[(11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2\right]$$

$$= 23$$

P

~~Step 3~~:- Covariance matrix $n \times n = 2 \times 2$

$$S = \begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix} = \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}$$

Step 4: Eigen value, Eigen vector, Normalize eigen vector.

→ Eigen value

$$\det(S - \lambda I) = 0$$

$$\det \begin{bmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{bmatrix} = 0$$

$$(14-\lambda)(23-\lambda) - (-11 \times -11) = 0$$

$$\lambda^2 - 37\lambda + 201 = 0$$

$$\lambda = 30.3849,\ 6.6151$$

$$\lambda_1 > \lambda_2$$

$$30.3849 > 6.6151$$

⇒ $\boxed{\text{First principal Component}}$

largest value

(ii) Eigen vector of λ_1

λ_1 = largest eigen value

u_1 = eigen vector of λ_1

$u_1 = 2 \times 1$

$$(S - \lambda_1 I) U_1 = 0$$

$$\begin{bmatrix} 14-\lambda_1 & -11 \\ -11 & 23-\lambda_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

$$\begin{bmatrix} (14-\lambda_1)u_1 - 11u_2 \\ -11u_1 + (23-\lambda_1)u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Linear eq.

$$(14-\lambda_1)u_1 - 11u_2 = 0$$

$$-11u_1 + (23-\lambda_1)u_2 = 0$$

Construct this linear eq:-

$$\frac{u_1}{11} = \frac{u_2}{14-\lambda_1} = t$$

when $t = 1$

$$u_1 = 11$$

$$u_2 = 14 - \lambda_1$$

eigen vector u_1 of $\lambda_1 = \begin{bmatrix} 11 \\ 14-\lambda_1 \end{bmatrix}$

$$= \begin{bmatrix} 11 \\ 14-30.3849 \end{bmatrix} = \begin{bmatrix} 11 \\ -16.38 \end{bmatrix}$$

Sign of Teacher

(iii) Normalize the eigen vector U,

$$e_1 = \begin{bmatrix} \frac{11}{\sqrt{11^2+(-16.38)^2}} \\ \frac{-16.3849}{\sqrt{11^2+(-16.38)^2}} \end{bmatrix} = \begin{bmatrix} .5574 \\ -0.8303 \end{bmatrix}$$

λ_2

$$e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}$$

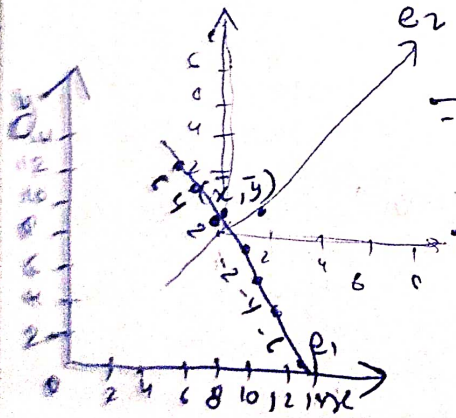
Step 5: Derive new dataset

$\boxed{P \cdot t_{01}}$

	Ex1	Ex2	Ex.3	Ex-31
FPC,	P_{11}	P_{12}	P_{13}	P_{14}
	-4.3052	3.7361	5.6928	-5.1238

e_1^T = Transpose

$$P_{11} = e_1^T \begin{bmatrix} 4-8 \\ 11-8.5 \end{bmatrix}$$



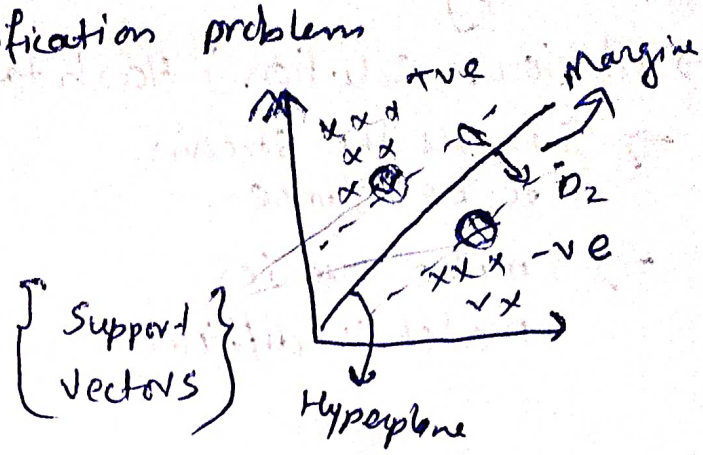
$$= \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} -4 \\ 2.5 \end{bmatrix}$$

$$= -4.3052$$

Support Vector machine

→ used for regression & classification problem

1. Support Vector
2. Hyperplane
3. Marginal Distance
4. Linear separable
5. Non linear separable



lower Dimension to higher dimension

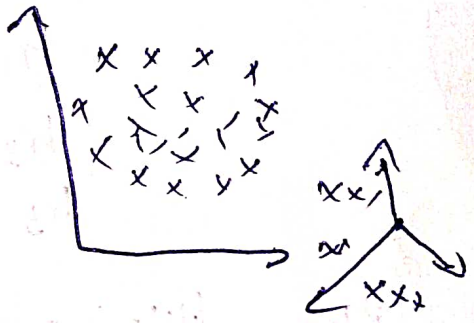
Sum kernel

2D

↓

High Dimension

D2 >> D1



Bayesian Data Analysis

1. Conditional Probability
2. Independent Event
3. Dependent Event

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

Step1:- Prior

Step2:- Parameter

Step3:- Generative model

Step4:- Stimulated data

Posterior

21%

4/16

Unit-II Data Analysis
- Univariate
- Bivariate
- Multivariate

Regression Modeling

→ A method to mathematically formulates relationship b/w variables
→ Analyze the relation b/w the target or dependent variable and independent variable in a dataset
→ This technique used to determine the time series, forecast trend.

Types of regression Modeling
- → linear Regression
- → Multiple Regression

- → linear Regression
- → Logistic Regression
- → Ridge Regression
- → Lasso Regression
- → Polynomial Regression
- → Bayesian linear Regression

M.L

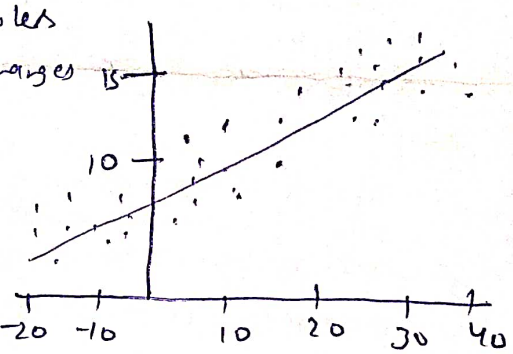
Linear Regression :-

→ basic type of regression
→ Consist :- predictor variable & a dependent Variable related linearly to each other
→ regression model equation :- $y = mx + c + e$

m :- Slope of line → have two variables one unit changes in result.
c = intercept
e = error in model

→ Mean square Prediction error
→ Quantity Prediction error.

Error = difference b/w the observed value & the predicted value.



The value of m & c get selected in such a way that it gives minimum predictor error.

→ Size of house & its selling price
→ KFC & no. of burger sold

* Univariate Analysis
(range, mean, mode, median)
Histogram, Piechart
Bar charts)

Multiple Regression

→ Consider more than one independent variable to predict a dependent variable.

* Bivariate Analysis
↳ Regression Analysis

Multivariate Analysis

→ To understand & reduce the data dimension by analyzing the data covariance structure.

→ deals with the statistical analysis of data collected on more than one dependent variable.

→ Factor Analysis, Cluster Analysis, Variance Analysis (techniques)

Key Concepts:-

1) The variate:- weighted sum of the variables.

The variate of n weighted variables (X_1 to X_n) written as

Variate:- $X_1 * W_1 + X_2 * W_2 + X_3 * W_3 + \dots + X_n * W_n$

Where $X_1, X_2 \dots X_n$ are observed variable

$W_1, W_2 \dots W_n$ are the weights

Techniques:-
→ overview
→ Initial step
→ multiple Regression Analysis
→ Logistic Regress
→ Discriminate Anal
→ Multivariate Analy
→ factor Anal
→ cluster Analys

2) Measurement Scale

data has multiple value (qualitative & quantitative), so to select appropriate multivariant technique, measurement is used.

Bayesian Modeling

R & Python are tools

→ It is when we use probability to represent uncertainity in all parts of a statistical model.

→ Probability of unknown quantities could be.

→ Bayesian inference is to learn from data

→

Bayes's rule:- $$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Inference & Bayesian Network:-

↳ probabilistic graphical model that uses Bayesian inference for probability computation.

↳ directed acyclic graph, in which edges are conditional dependency, nodes are random variables.

Conclusion

Inference can come in two forms:-

- simply evaluate the joint probability of a particular assignment of values for each variable in the network.
- second to find $P(X|e)$, or to find the probability of some assignment.

Filtering Streams :-

Webcrawler :- Search url

Url Hyperlink

→ obvious solution : Hash table

⇒ S is all the stream to be stove

⇒ 1 million queries match is difficut

Put filtering → DB include

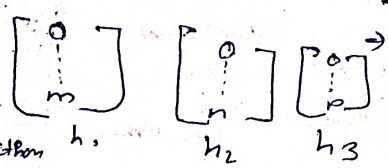
match → x
not match - added

email
Inbox spam
unwanted new word

Emailed → 20 bytes

→ queries ⇒ store cache [0 . . . m]
m bit

Hashmap
↓
bit array

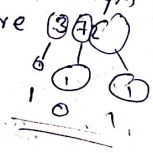


Hash function h_1 h_2 h_3

word = "abc"

h_1 (abc) = 3
h_2 (abc) = 7
h_3 (abc) = 2

We will not store 20 bytes, we will store



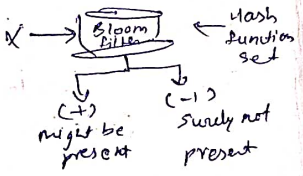
Bloom filtering :-

~~Set~~ ~~stream~~ → n bit array.
→ N = 11 00000000000
→ Stream element = integer

$h_1(x)$ = add ~~dig~~ places digits
$h_2(x)$ = even places digit

used to detect whether a element is present in the pass or not

It is FALSE Positive but not false negative



Ex :-
A set consist of some element say, 8, 10 .... check whether 48, 7 lie in the set or Not.

Set ~~Arr~~ size = 10 (Array size)

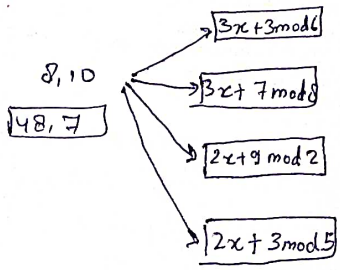
Hash function set

- $(3x + 3) \bmod 6$
- $(3x + 7) \bmod 8$
- $(2x + 9) \bmod 2$
- $(2x + 3) \bmod 5$

Step 1 :- Initilize Array

0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0

Step 2 :- Apply Hash function & Calculate the value.



8	10	48 (random)	7
3	3	2	0
1	5	7	
7	1		
4	3		

0	1	2	3	4	5	6	7	8	9
0	1	0	1	1	1	0	1	0	0

Step 3 :-
all array values are - 1 then False present
all array value are not 1 then True not present

Regression Equation

Q. From Given data, Calculate regression eq. taking deviation of items from the mean of X & Y Series.

X	1	2	3	4	5
Y	2	5	11	8	14

Solution: Types of regression equation

info (1) Regression eqn of y on X
(2) " " " X on Y

- Y on X → variation in Y when the changes in X is given
- X on Y → variation in X when the changes in Y is given.

- Regression eq. of X on Y

$$(X - \bar{X}) = \frac{r.\Sigma x}{\Sigma y}(y - \bar{y})$$

↳ Regression Co-efficient.

$$b_{xy} = \frac{r.\Sigma x}{\Sigma y} = \frac{\Sigma xy}{\Sigma y^2}$$

- Regression eq. of Y on X

$$(Y - \bar{Y}) = \frac{r\,\Sigma y}{\Sigma x}(X - \bar{X})$$

↳ Regression Co-efficient

$$b_{yx} = \frac{r\,\Sigma y}{\Sigma x} = \frac{\Sigma xy}{\Sigma x^2}$$

r for relational Co-efficient

$$r = \sqrt{b_{xy} \cdot b_{yx}}$$

Solution:- Q-2

X	$x=(X-\bar{X})$	x^2	Y	$y=(Y-\bar{Y})$	y^2	XY
1	-2	4	2	-6	36	12
2	-1	1	5	-3	9	3
3	0	0	11	3	9	0
4	1	1	8	0	0	0
5	2	4	14	6	36	12
$\Sigma X=15$		$\Sigma x^2=10$	$\Sigma Y=40$		Σy^2	$\Sigma xy=27$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{15}{5} = 3$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{40}{5} = 8$$

- Regression equation of X on Y

$$(X-\bar{X}) = \frac{\Sigma xy}{\Sigma y^2}(Y-\bar{Y})$$

$$(x-3) = \frac{27}{90}(Y-8)$$

$$x-3 = 0.3y - 2.4$$

$$x - 0.3y - 0.6 = 0$$

$$\boxed{x = 0.3y + 0.6}$$

- Regression eq of Y on X

$$y = 2.7x - 0.1$$

Correlation Cofficient:

$$r = \sqrt{b_{xy} \cdot b_{yx}}$$

$$= \sqrt{\frac{\Sigma xy}{\Sigma y^2} \times \frac{\Sigma xy}{\Sigma x^2}} = \sqrt{2.7 \times 0.3}$$

$$= 0.9$$

Answer

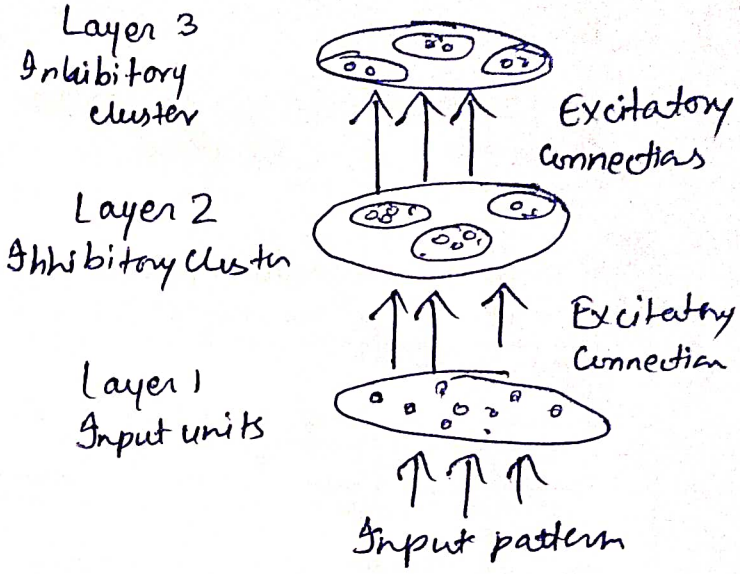
Competitive learning :-

→ ~~A set of neurons that are all same except for some randomly distributed synaptic weights.~~

→ type of unsupervised learning model used in machine learning and artificial intelligence system.

→ It consist of a set of hierarchically layered units in which each layer connects.

→ Competitive approach means Student work separately on a project



Fuzzy logic :- (Zadeh, Zadeh)

→ Application :- Dishwasher, Copymachine
Aerospace, chemicals,
Automobile System (gear system)

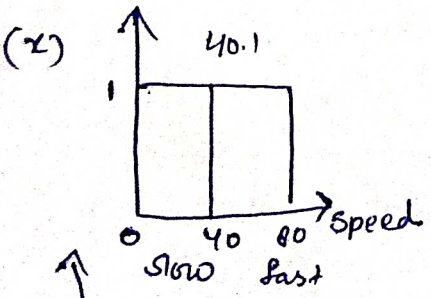
Boolean logic	Vs	fuzzy logic
is it hot → Yes/1, No/0		Is it hot → Very much/0, fairly (.7), moderate (.5), somewhat (.25), very little (.1)

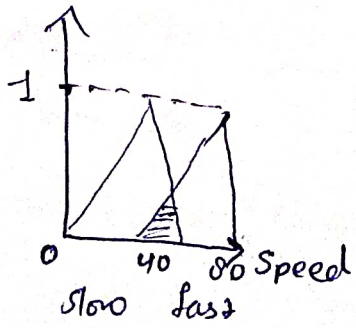
- Extracting fuzzy models from the data.

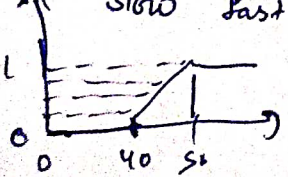
→ It means some fuzzy rules from data allow relationship in the data to be modeled by "if-then" rules that are easy to understand, verify, & extend.

uncertainity

membership function







Check the degree of fastness

$$\begin{cases} 0, & \text{if } Speed(x) \le 40 \\ \frac{Speed(x) - 40}{10}, & \text{if } 40 < Speed(x) < 50 \\ 1, & \text{if } Speed(x) \ge 50 \end{cases}$$

$x = 30 = \{30, 0\}$ $x = 42$ (.2
$x = 60 = \{60, 1\}$ $x = 45$ (.5

Ex

$U = \{1, 2, 3, 4, 5\}$

$S = \{1, 2\}$

$= \{(1,1)\ (2,1)\ (3,0), (4,0)\ (5,0)\}$

x, m

* Fuzzy decision tree

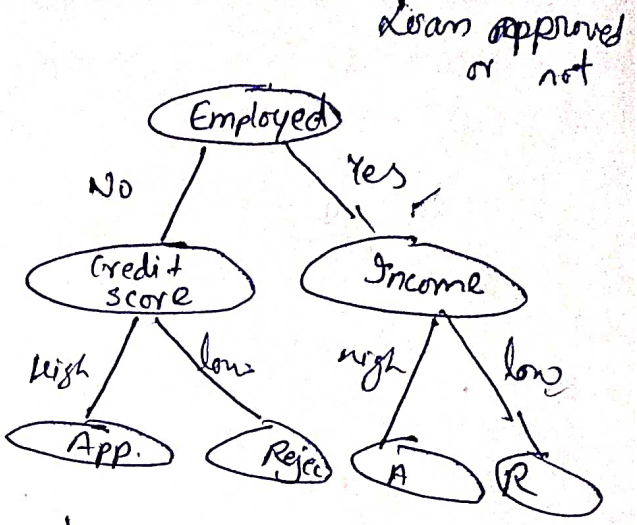
→ A classifier

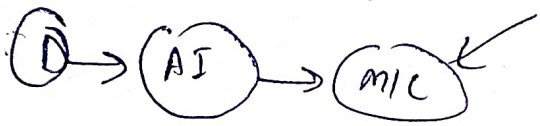
→ Also used for regression

→ Decision nodes (test)

→ Leaf node (classification/value)

→ Test is performed on the feature.





Employee	Credit score	Income
Y		
Y		
Y		
N		
N		

* Stochastic Search methods

→ Simulated Annealing (SA

→ Probabilistic Global search Lausanne (PGSL)

→ Genetic algorithm (GA)

- It refers to the use of randomness in the objective function or in optimization algorithm.

→ Ex:- → arranging molecules as regular, crystal structures at appropriate temperature reduction.

→ Creative adaptive, learning organism through biological evolution.