CS349

Artificial Intelligence-2

UNCERTANITY :

→ Quantify Uncertanity → for every event in a series of events we need to quantify.

(Partial Observation)
→ Agent may need to handle whether due to P.O., non-determinism, or combining these the agent may never know for certain what state it is in or where it will ends up after a sequence of actions

→ Problem solving Agents } AI-1
→ LOGICAL Agents

designed to handle uncertainity by keeping track of a BELIEF STATE (A representation of all possible world states)

HANDLING UNCERTAIN KNOWLEDGE :

→ $\forall p$ Symptom (p, toothache) → Disease (p, Cavity)
✗ Problem

not correct : T.A can be caused by many other causes
T.A → toothache

→ $\forall$ P Symptom (P, toothache) → D (P, Cavity) ∨ D(P, Gumdisease) ∨ D(P, wisdom teeth) ∨ - - . . . - -
D → Disease

→ $\forall$ P Disease (P, cavity) → Symptom (P, Toothache)
✗

not correct : all cavities don't cause tooth ache

REASONS TO USE PROBABILITIES:

1. specifications come too large: It is too much work to list the complete set of antecedents or consequents needed to assume an exceptionless rule. (Laziness)

2. THEORITICAL IGNORANCE : Complete set of antecedents are not known

3. Practical Ignorance

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

→ Probability that x is fat = 0.2

if x is fat then x has coronary heart disease (CHD) with prob = 0.7

$P(x \text{ has CHD}) \approx (0.2 \times 0.7) + (*)$

→ Data ↑ ⇒ Knowledge ↑ ⇒ certainity ↑

AXIOMS OF PROBABILITY:

1) All probability btw. 0 and 1.

$$0 \leq P(A) \leq 1$$

2) $P(\text{True}) = 1$, $P(\text{False}) = 0$

3) $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

4) Bayes : $P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$ $(P(A) \neq 0)$

$\to P(A \cap B)$

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

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

→ we use probability to quantify the uncertainity.

→ Decision making function agent, returns all actions

Persistent : belief state (Prob. beliefs about the current state of the world)

update belief state based on Action and percept

calculate outcome probability for Actions

↳ given action description and belief state

select action with highest expected Utility

given prob. of outcome and Utility Info

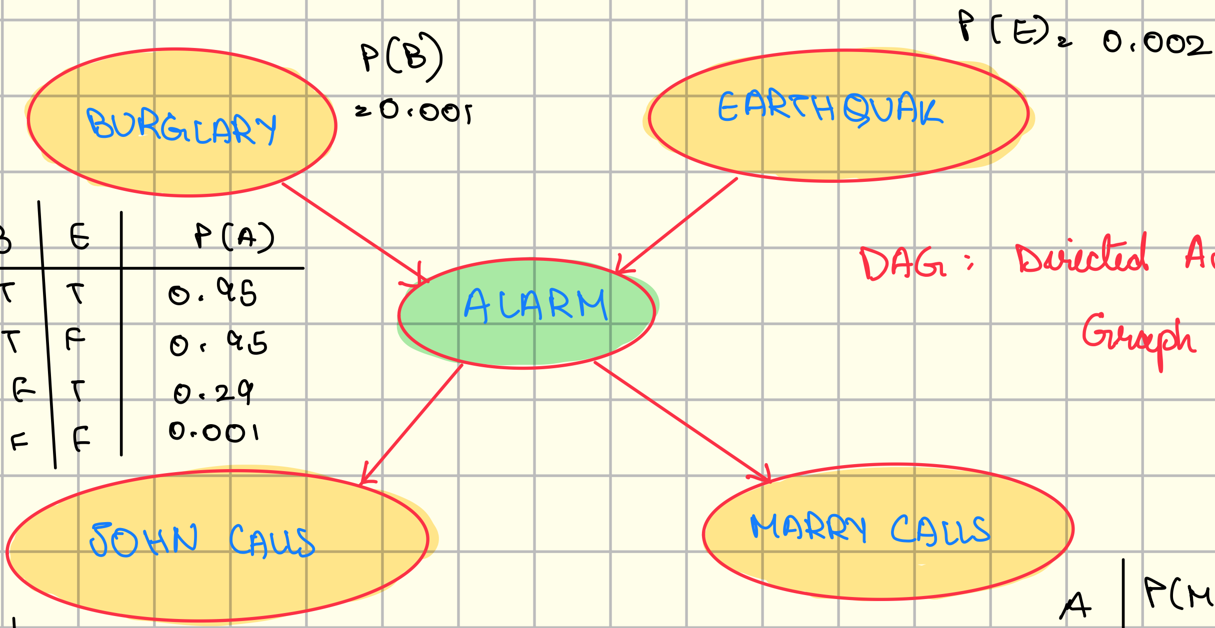
Return Action

→ Decision Theory = Prob. theory + Utility Theory

→ The Fundamental idea of DT is that an Agent is Rational iff it chooses the actions that yields the highest expected Utility, averaged over all possible outcomes of Action

↳ Maximum Expected Utility (MEU)

BELIEF NETWORK EX. :



DAG : Directed Acyclic Graph

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

A	P(J)
T	0.9
F	0.05

A	P(M)
T	0.7
F	0.01

→ A Belief Network is a graph with following :

① Nodes : set of Random variables

② Directed lines : The Intutive meaning of an edge from node X to node Y is that X has direct influence on node Y

③ Each node has a conditional Probability table that quantifies the effect that the Parent have on the node.

④ DAG (no Cycles)

→ BUGLAR ALARM at home:

↳ fairly reliable on detecting BUGLAR

↳ Responds at times to minor EQ

→ Two neighbour on hearing ALARM calls police

↳ John always calls when he hears the alarm

↳ But sometimes confuses the telephone ringing with the ALARM & calls them too.

↳ MARRY likes LOUD MUSIC and sometimes misses the ALARM

The Joint Prob. dist.:

→ a generic entry in the Joint Prob. distribution is given by

↳ $P(x_1, x_2, \ldots, x_n)$

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{Parents}(x_i))$$

→ Prob. of the event that the ALARM has sounded but neigther a Buglar nor a EQ has occured, but both John & Marry call.

→ $P(J \wedge M \wedge A \wedge \bar{B} \wedge \bar{E}) = P(J \mid A)\, P(M \mid A)\, P(A \mid \bar{B} \wedge \bar{E})\, P(\bar{B})\, P(\bar{E})$

$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$

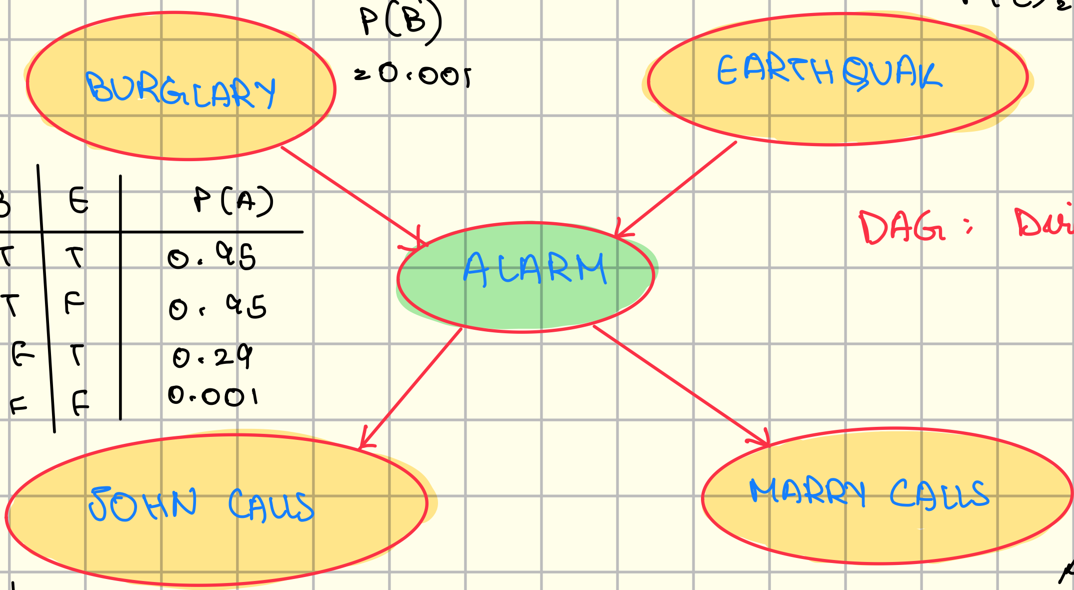
Joint Prob. Distribution :

→ A generic entry in the join prob. dist. is given by,

$$P(x_1, x_2 \dots x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(x_i))$$

$P(E) = 0.002$

$P(B) = 0.001$



DAG : Directed Acyclic Graph

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

A	P(J)
T	0.9
F	0.05

A	P(M)
T	0.7
F	0.01

$$P(J) = P(J|A)P(A) + P(J|A')P(A')$$

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

$$= 0.00062$$

$$P(A) = P(AB'E') + P(AB'E) + P(ABE') + P(ABE)$$

$$= P(A|B'E')P(B'E') + P(A|B'E)P(B'E) + P(A|BE')P(BE') + P(A|BE)P(BE)$$

$$= 0.001 \times 0.999 \times 0.998 + 0.29 \times 0.999 \times 0.002 +$$

$$0.95 \times 0.001 \times 0.998 + 0.95 \times 0.001 \times 0.002$$

$$= 0.0025$$

$P(J) = P(JA) + P(JA') = P(J|A)P(A) + P(J|A')P(A)$

$= 0.9 \times 0.0025 + 0.05 \times (1-0.0025)$

$= 0.052125$

$P(AB) = P(ABE) + P(ABE') = P(A|BE)P(BE) + P(A|BE')P(BE')$

$= 0.95 \times 0.001 \times 0.002 + 0.95 \times 0.001 \times 0.998$

$= 0.00095$

$P(A'B) = 0.00005$ $\quad P(A'E') = 0.996$

$P(AE) = 0.00058$ $\quad P(JB) = 0.00086$

$P(AE') = 0.001945$ $\quad P(B) = 0.001$

$P(J|B) = \dfrac{P(JB)}{P(B)} = \dfrac{0.00086}{0.001} =$

$P(MB) = 0.00067$

$P(M|B) = \dfrac{P(MB)}{P(B)}$

$P(MB) = 0.00067$

$P(M|B) = 0.67$

$P(B|J) = 0.016$

$P(B|A) = 0.3$

$P(B|AE) = P(ABE) | P(AE) = \dfrac{P(A|BE)P(BE)}{P(AE)} = \dfrac{0.95 \times 0.001 \times 0.002}{0.00058}$

$= 0.003$

$P(JE') = P(AJE') + P(A'JE')$

$P(A'JE')$

$P(BJE')$

$$P(B|JE') = \frac{P(BJE')}{P(JE')}$$

INFERENCE USING BELIEF NETWORK:

→ DIAGNOSTIC INFERENCES (from effect to causes)

given that John calls, infer that

$P(\text{Buglery} \mid \text{John calls}) = 0.016$

→ CAUSAL INFERENCE

given Buglary, infer that

$P(\text{John calls} \mid \text{Buglary}) = 0.86$

$P(\text{Mary calls} \mid \text{Buglary}) = 0.67$

→ INTER-CAUSAL INFERENCES (btw. causes & common effect)

Given A, we have $P(B|A) = 0.376$

if we add evidence that E is true, then $P(B|A \wedge E) = 0.003$

→ MIXED INFERENCE:

Setting that John calls to TRUE and the CAUSE E.Q. to false.

$P(A \mid \delta \wedge \bar{E}) = 0.003$

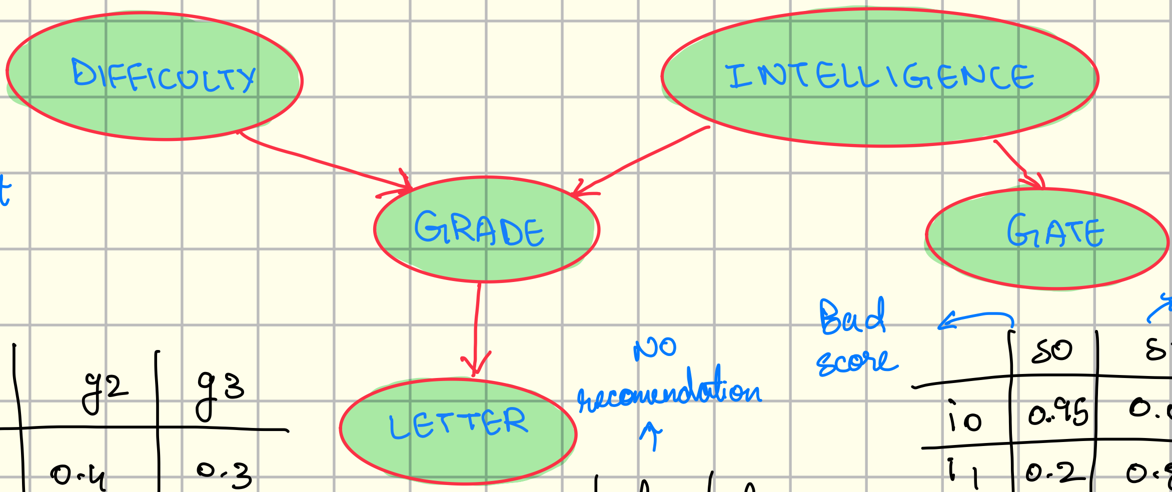
Not intelligent

I_0	I_1
0.7	0.3

d_0	d_1
0.6	0.4

not difficult paper

difficult



	g_1	g_2	g_3
i_0, d_0	0.3	0.4	0.3
i_0, d_1	0.5	0.25	0.7
i_1, d_0	0.9	0.08	0.02
i_1, d_1	0.5	0.3	0.2

No recomendation

	l_0	l_1
g_1	0.1	0.9
g_2	0.4	0.6
g_3	0.99	0.01

Bad score

good score

	s_0	s_1
i_0	0.95	0.05
i_1	0.2	0.8

Q1) what is the prob. of getting excellent recconendation letter given the Question very easy.

Q2) what is the Prob. of the student being very intelligente gets a poor Recomendation letter?

Q3) what is the Probability of getting poor grade score given grade good?

Conditional Independence:

$$P(x_1, \ldots x_n) = P(x_n \mid x_{n-1}, \ldots x_1)\, P(x_{n-1}, \ldots x_1)$$

$$= P(x_n \mid x_{n-1}, \ldots x_1)\, P(x_{n-1} \mid x_{n-2} \ldots x_1) \ldots P(x_2 \mid x_1) P(x_1)$$

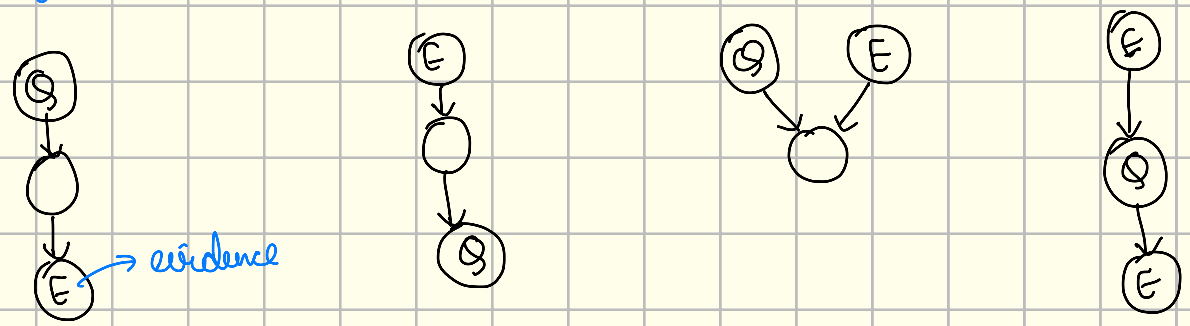
$$= \prod_{i=1}^{n} P(x_i \mid x_{i-1}, \ldots, x_1)$$

→ The belief network represents conditional Prob.

$$P(x_i \mid x_{i-1} \ldots x_1) = P(x_i \mid Parents(x_i))$$

Graph Paterns

Diagnostic | Causal | Inter Causal | Mixed



$P(Q \mid E)$

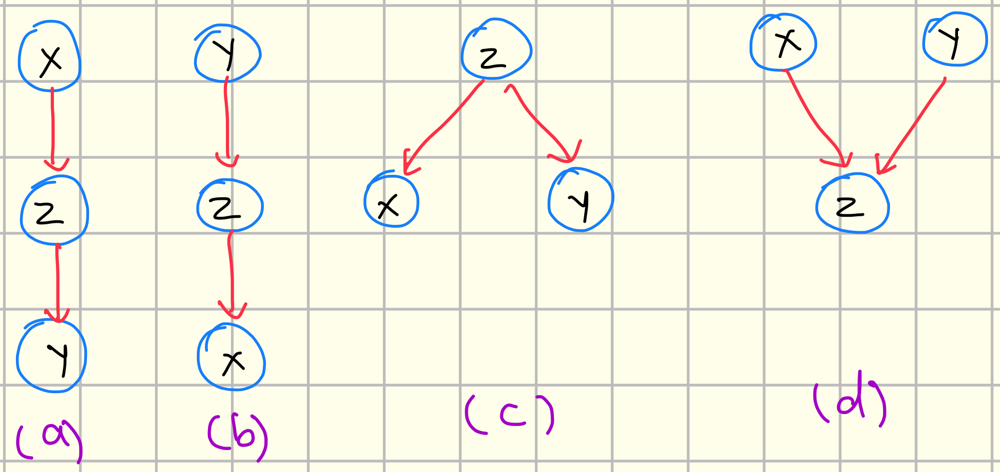
Incremental Network Construction :

1) Choose the set of relavent x_i that describes the domain

2) Choose an ORDERING for the variables

3) while there are variables left
    a) Pick variable x and add a node for it
    b) Set Parent (x) to some minimal set of existing nodes such that conditional independence property is satisfied
    c) Define cond^n Prob. tables for x.

Conditional Independence Relations



→ A path is blocked given a set of nodes E if there is a node z on the path for which one of 3 conditions holds.

1) z is in E and z has one arrow on the path leading in and 1 arrow out (a, b) X & Y are independent given Z, X & Y are indepen-

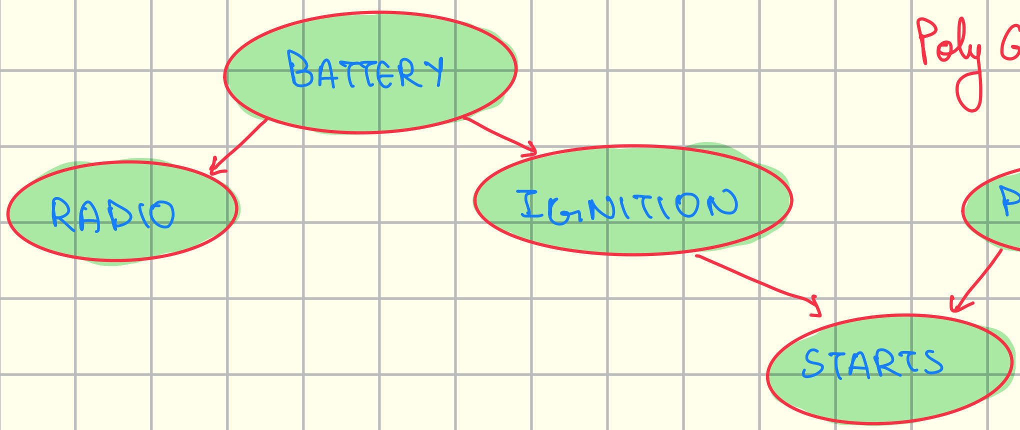
2) Z is in E and z has both path arrows leading out (c)

3) Neigther Z nor any decendent of z is in E, and both path arrows lead to z (d) if Z and its succesors are not given ∴ X & Y are independent

→ if every undirected path from a node in X to a node in Y is deseperated by a given set of evidence nodes E then X & Y are Conditionally indipendent given E

→ A set of nodes E desperated two sets of nodes X & Y if every undirected path from a node in X to a node in Y is blocked given E.

Poly Graphs



(1) → whether there is petrol and whether the radio plays are independent given evidence about whether ignition takes place

(2) → R and P are independent if it is known B works
evidence

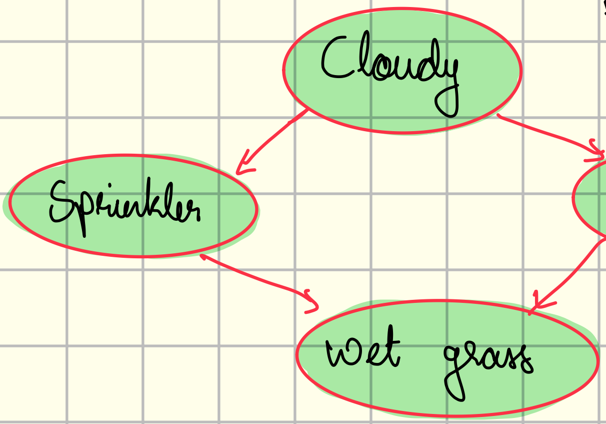
(3) → Petrol and R are independent given no evidence at all

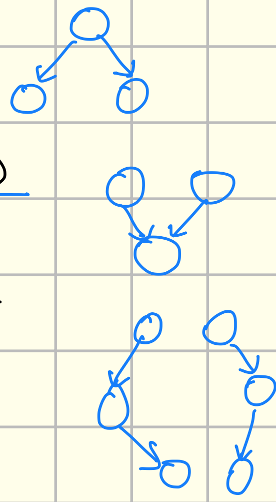
(4) → P and R are dependent given Evidence whether car starts

(5) → (If car does not start, then R playing is evidence that out of petrol.)

→ Inference is multiply connected Belief Network

$P(C) = 0.5$





C	P(S)
T	0.1
F	0.5

C	P(R)
T	0.8
F	0.2

S	R	P(w)
T	T	0.99
T	F	0.9
F	T	0.9
F	F	0

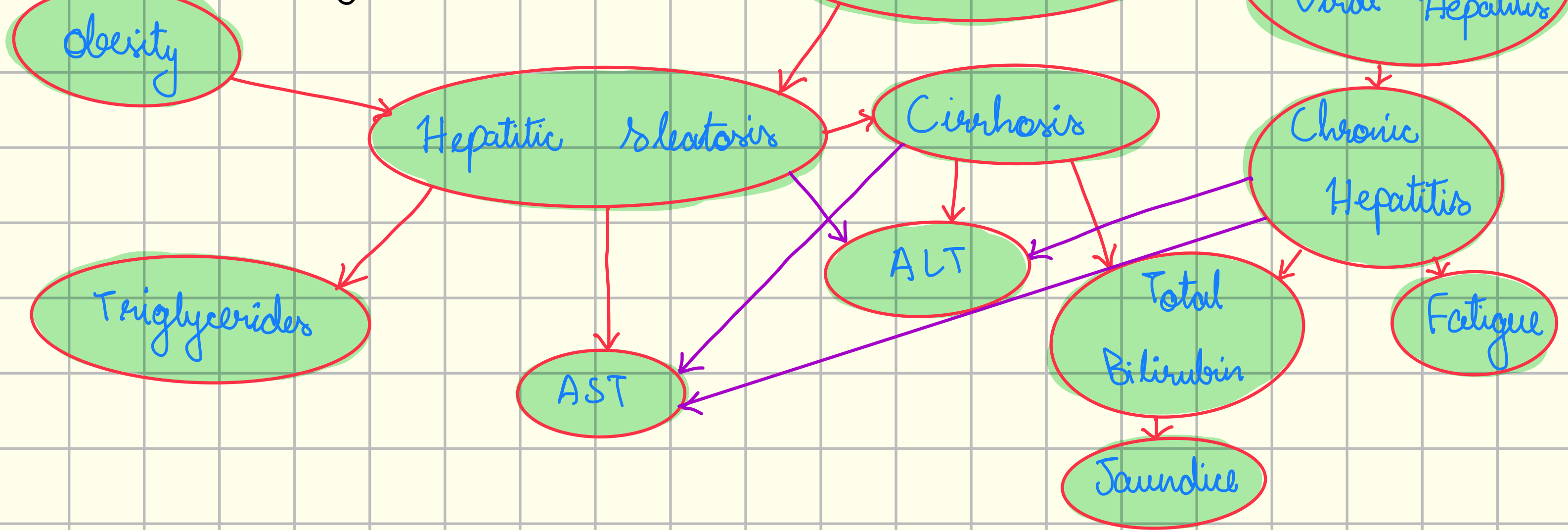
→ (knowing w) S & R are not ind.

→ (knowing c) S & R are ind.

Bayesian Belief Network :

T 0.3
F 0.05

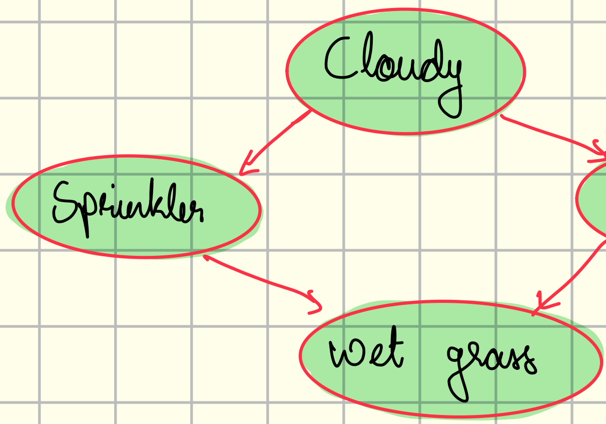
P(obesity) = 0.24



AST : Asparate Amino Transference

ALT : Alanino " "

$P(C) = 0.5$



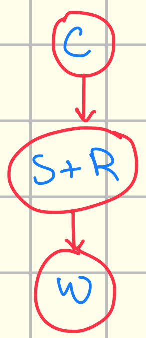
C	P(S)
T	0.1
F	0.5

C	P(R)
T	0.8
F	0.2

S	R	P(W)
T	T	0.99
T	F	0.9
F	T	0.9
F	F	0

Clustering Method:

→ transform the network into a probabilistically equivalent (but topologically different) Poly tree by merging ordering nodes

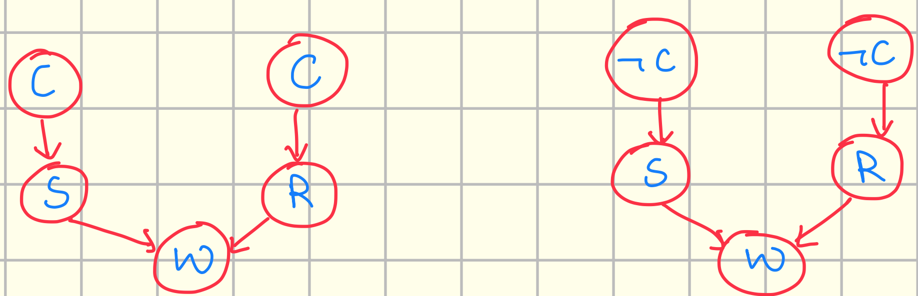


$P(C) = 0.5$

C	TT	TF	FT	FF
T	0.08	0.02	0.72	0.18
F	0.4	0.1	0.4	0.1

S	R	P(W)
T	T	0.99
T	F	0.9
F	T	0.9
F	F	0

Cut Set Conditioning Method:



→ A set of variables that can be instantiated to yield a Poly-tree is called cutset

→ Instantiated the cutset variable to define values then evaluate a polytree for each possible instantiation.

Inference in Multiple connected Belief Network

→ Stochastic Simulation Method

→ Use the N/w to generate a large no. of concrete models of the domain that are consistent with the Network distribution

→ They given an approx. of exact evaluation

Simpson's Paradox :

Male	Recovered	Not Recover	Recovery Rate
Given drug	18	12	60 %
Not given drug	7	13	35 %

Female	Recovery	Not Recover	Recovery Rate
Given drug	2	8	20 %
Not given drug	9	21	30 %

Combined	Recovery	Not Recover	Recovery Rate
given drug	20	20	50%
Not given drug	16	24	40%

→ Should the drug be administered or not ?

$P(Rec \mid M \wedge D) = 0.6$

$P(R \mid F \wedge D) = 0.2$

$P(R \mid \text{given drug}) = P(R \mid M \wedge D)\, P(D \mid M) + P(R \mid F \wedge D)\, P(D \mid F)$

$= 0.6 \times \frac{30}{40} + 0.2 \times \frac{10}{40} = 0.5$

Dampster - Shafer Theory

→ Designed to deal with the distribution btw UNCERTAINITY & IGNORANCE

→ we use a belief function $B(x)$: Probability that the Evidence supports the preposition

→ when we don't have any evidence about x we assign $B(x) = 0$ as well as $B(\bar{x}) = 0$

→ if we are given that the coin is fair with 90% certainity then

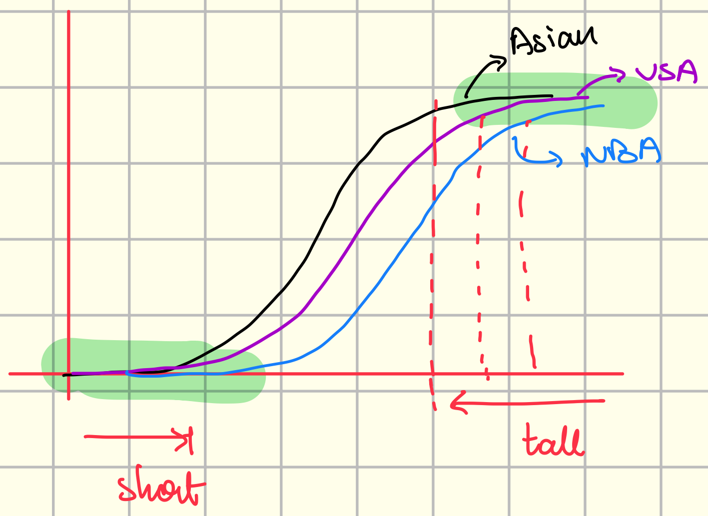
$B(\text{head}) = 0.9 \times 0.5 = 0.45$

$B(\overline{\text{head}}) = 0.9 \times 0.5 = 0.45$

→ we will have a gap of 0.1 that is not accounted by the given evidence

FUZZY LOGIC :

→ Paradigm shift from Probability to Possibility



→ Probability of an event → Possibility (degree of truth of an event)

either fat or non-fat (Boolean)

If is not Boolean, it is a set {very fat, medium fat, ------}

if very fat more chances of cardiac problem

" less " less " " " "

" NO " NO " " " "

(Possibility of cardiac problem depends on how fat she is.)

→ $T(A \wedge B) = \min(T(A), T(B))$. A & B are 2 events

$T(A \vee B) = \max(T(A), T(B))$

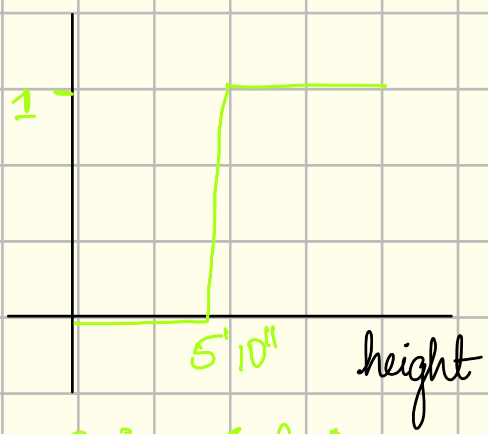
Fuzzy logic

Fuzzy set theory is a mean of specifying how will an object satisfies a vague description.

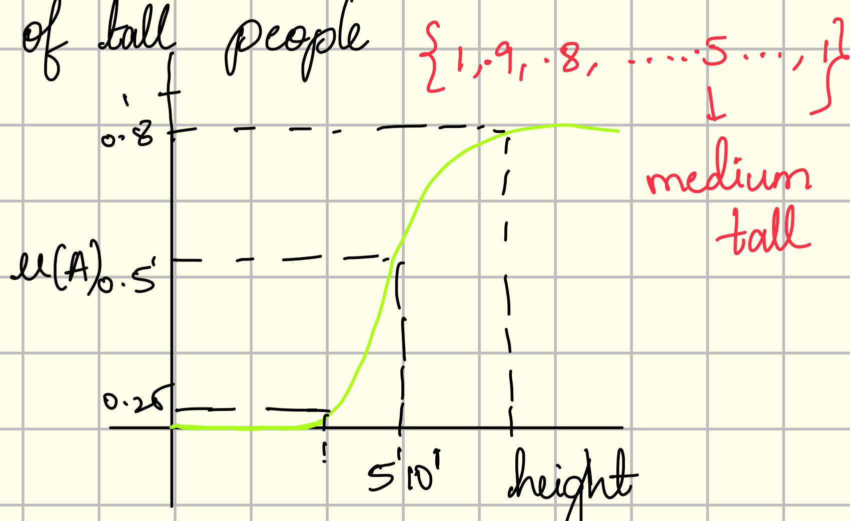
- Truth is a value between 0 and 1
- Uncertainty means from lack of evidence. but given the dimensions of a man whether he is fat her no uncertainty involves

Set with fuzzy Boundaries.

A = set of tall people



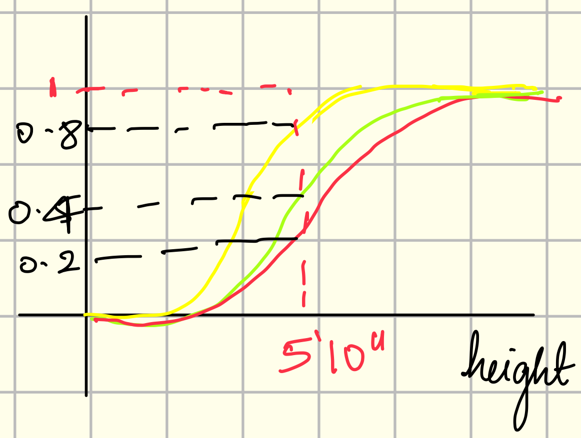
Crisp Set A



Fuzzy set A

Membership function (MF):

Subjective measure not probability funct^n



A Fuzzy set A in X is expression as a set of ordered pair.

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

universe or

Fuzzy Set

Membership funct^n (MF)

Uni of discourse

A F. Set is totally characterized by a MF

* Fuzzy set with discreate universe

Fuzzy set C = "desirable city to live in"

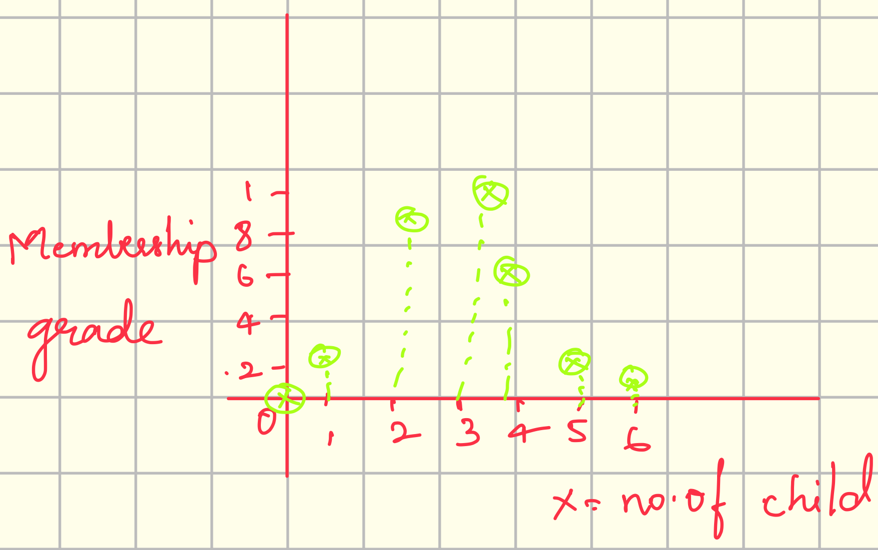
X = { Bihta, MP, SM } (discrete and non-ordered)

C = { (Bihta, 0.9), (MP, 0.8), (SM, 0.6) }

Fuzzy set A = "Sensible no. of children"

X = { 0, 1, 2, 3, 4, 5, 6 } (discrete Uni)

A = { (0, 0.1), (1, 0.3), (2, 0.7), (3, 1), (4, 0.6), (5, 0.2), (6, 0.1) }



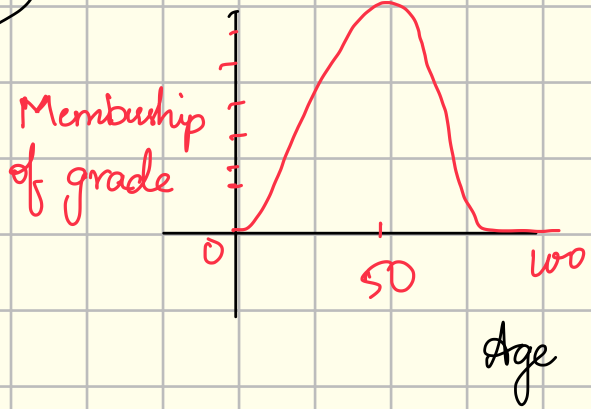
Fuzzy Set with continous Universe

fuzzy Set B = "about 50 years old"

X = Set of positive real values ℝ (Continous)

$$B = \{(x, \mu_B(x)) \mid x \in X\}$$

$$\mu_B(x) = \frac{1}{1+\left(\frac{x-50}{10}\right)^2}$$



Fuzzy Rules:

→ Diet is low and exercise is high ⇒ Balanced

→ Diet is high or exercise is low ⇒ Unbalanced

→ Balanced ⇒ Risk is low

→ Unbalanced ⇒ Risk is high

→ For a person it is given that
exercise = Burning 1000 calories/day

What is the risk of heart disease?

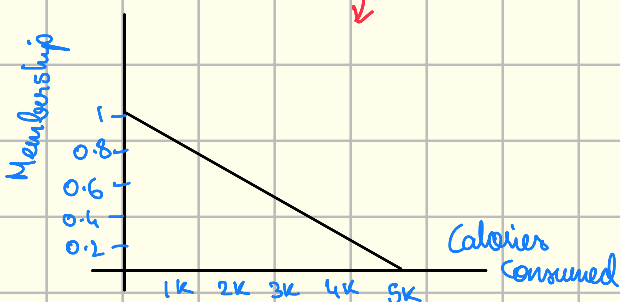
Membership Function (MF)

$$f_{DH}(x) = \frac{x}{5000}$$

(Diet high)

$$f_{DL}(x) = 1 - \frac{x}{5000}$$

(diet low)



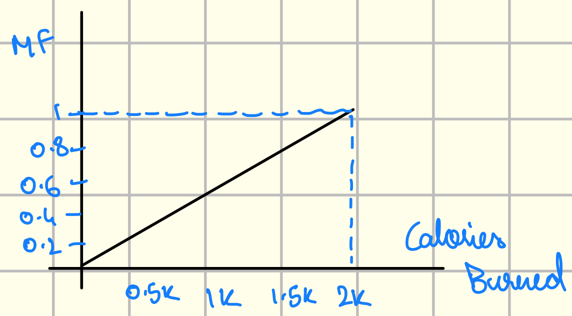
→ for daily calorie intake of 3000

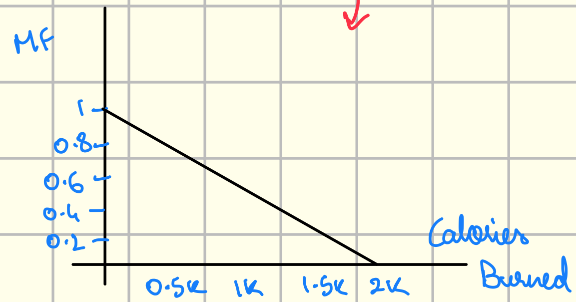
Membership of DH $= \frac{3000}{5000} = 0.6$

" " DL $= 0.4$

→ $f_{EH} = \frac{x}{2000}$ (Exercise high)

$f_{EL} = 1 - \frac{x}{2000}$ (Exercise Low)





→ for daily burnout of 1K calories:

Membership of EH $= 0.5$

" " EL $= 0.5$

Rule Evaluation:

→ truth (Diet high) $= T(DH) = 0.6$ | $T(DL) = 0.4$

$T(EH) = 0.5$ | $T(EL) = 0.5$

→ $DL \wedge EH \Rightarrow$ Balanced

$T(\text{Balanced}) = \min\{T(DL), T(EH)\}$

$= \min(0.4, 0.5)$

$= 0.4$

→ $DH \vee EL \Rightarrow$ Unbalanced

$T(\text{Unbalanced}) = \max(T(DH), T(EL))$

$= \max(0.6, 0.5)$

$= 0.6$

→ Balance ⇒ Risk ↓

Unbalance ⇒ Risk ↑

$T(R\downarrow) = T(Bal) = 0.4$

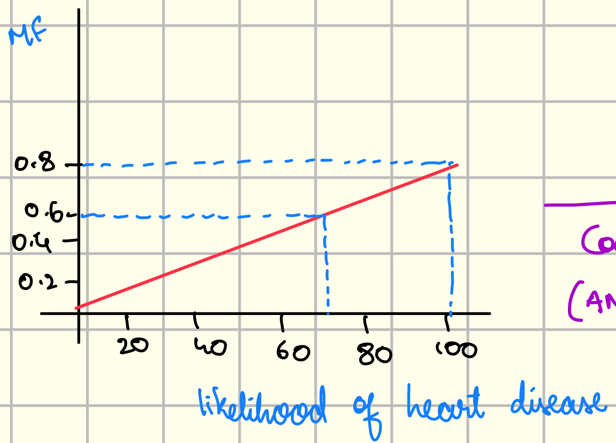
$T(R\uparrow) = T(UnBal) = 0.6$

→ $f_{R\uparrow} = f_{RH} = \dfrac{x}{125}$
(Risk high)

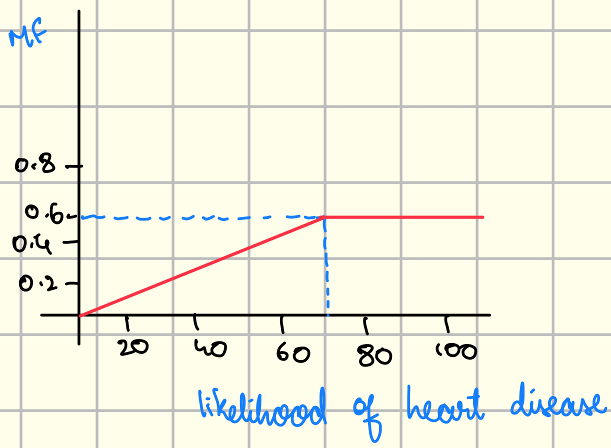
$T(R\uparrow) = T(RH) = 0.6$

$0.6 = \dfrac{x}{125}$

$\Rightarrow x = 75$



Combining (AND == min)

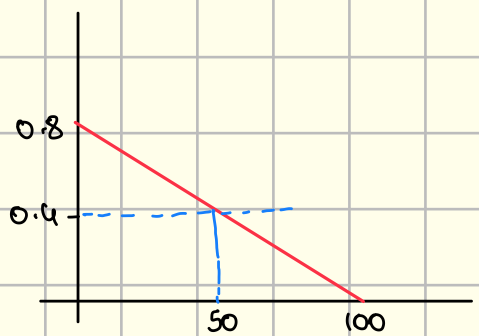


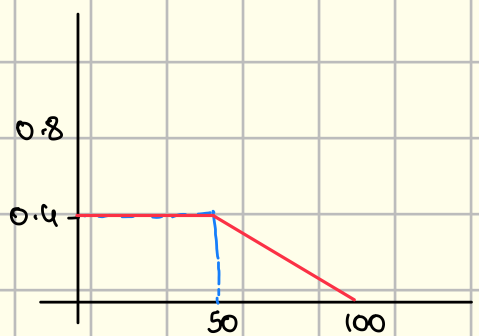
→ $f_{R\downarrow} = f_{RL} = 0.8 - \dfrac{x}{125}$

$T(RL) = 0.4$

$0.4 = 0.8 - \dfrac{x}{125}$

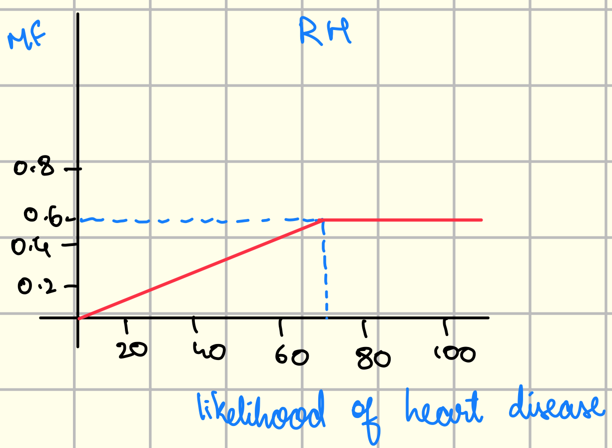
$\Rightarrow x = 50$

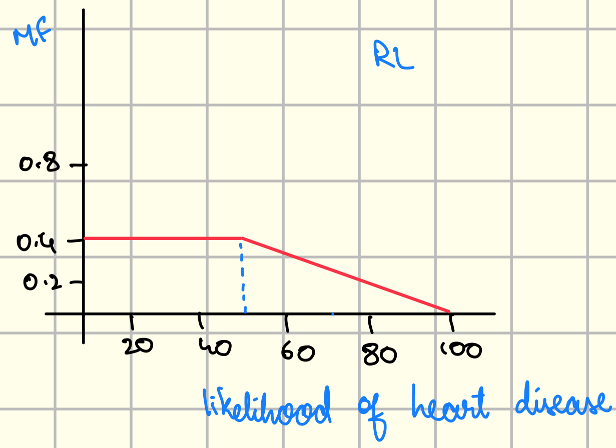




→ $f_{\substack{\text{aggregated}\\ \text{Risk}}}(x) = \begin{cases} 0.4, & \text{if } x \in [0, 50) \\ 3x+1, & \text{if } x \in [50, 75) \\ 0.6, & \text{if } x \in [75, 100] \end{cases}$

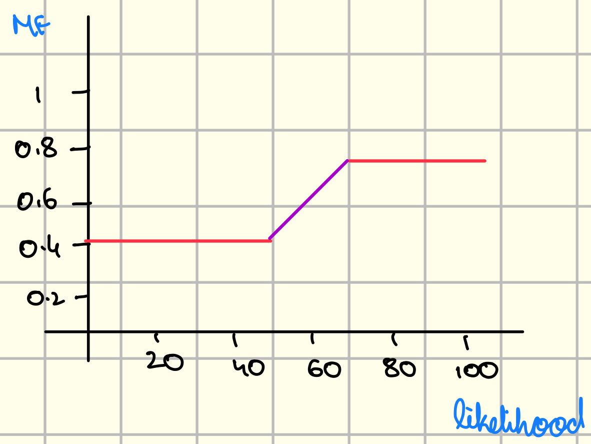
(f_{AH})





$T(A \cap B) = \min(T(A), T(B))$

$T(A \cup B) = \max(T(A), T(B))$



→ Defuzzification

$$\int_0^{100} f_{AH} \cdot dx = \int_0^{50} 0.4\, dx + \int_{50}^{75} \frac{x}{125}\, dx + \int_{75}^{100} 0.8\, dx$$

$$= 50 \times 0.4 + \frac{1}{125}\left[\frac{x^2}{2}\right]_{50}^{75} + 25 \times 0.8$$

$$= 47.5$$

→ the likelihood of HD for the person is 47.5%
(Heart Disease)

Altermeier's Notion of Fuzzy sets in Continuous & discrete

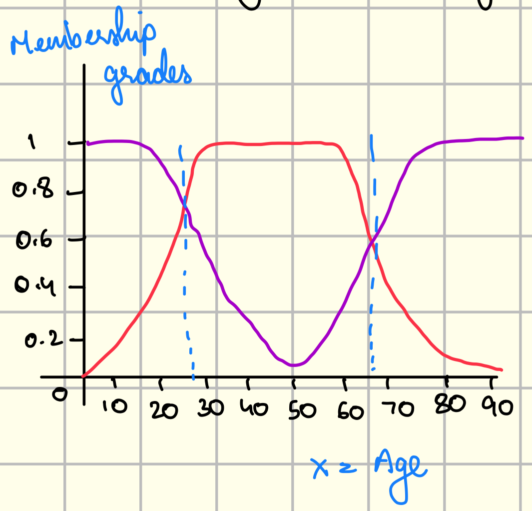
→ x is discrete $\Rightarrow$ $A = \sum_{x_i \in x} \mu_A(x_i) | x_i$

→ x is continuous ⇒ $A = \int_x \mu_A(x) / x$

"/" stands for a marker, not decision

FUZZY PARTITIONS :

→ formed by the linguistic values, "Young", "Middle Age", "old"

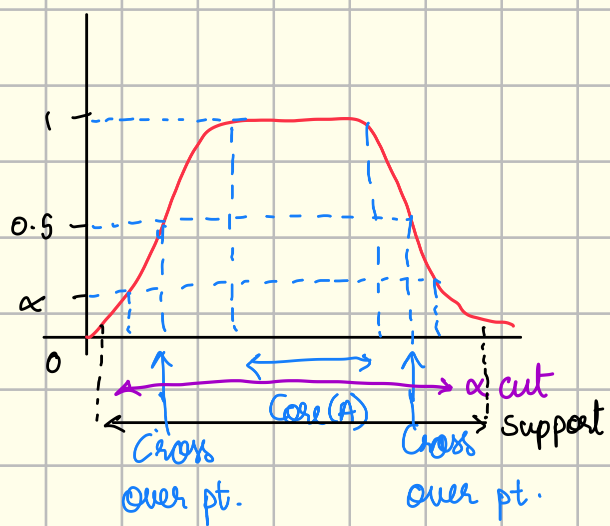


BASIC DEFINITIONS & TERMINOLOGIES

→ $\text{Support}(A) = \{x \in X \mid \mu_A(x) > 0\}$

$\text{Core}(A) = \{x \in X \mid \mu_A(x) = 1\}$

NORMALITY : $\text{Core}(A) \neq \phi$ [A is a normal fuzzy set]



$\text{crossover}(A) = \{x = X \mid \mu_A(x) = 0.5\}$

α-cut : $A_\alpha = \{x = X \mid \mu_A(x) \geq \alpha\}$

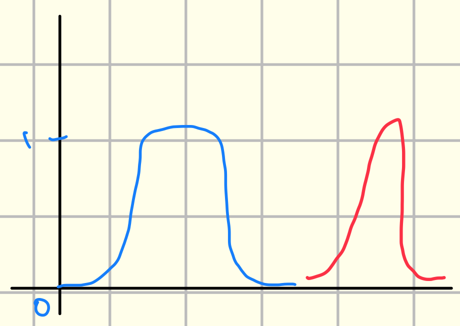
strong α-cut : $A'_\alpha = \{x = X \mid \mu_A(x) > \alpha\}$

Convexity of Fuzzy Set :

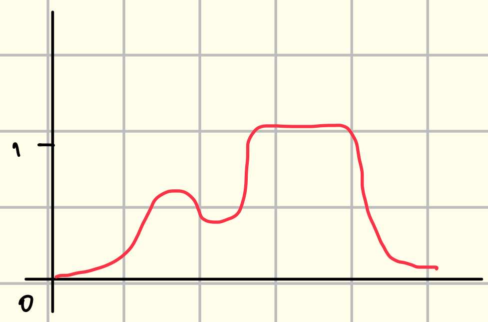
→ A fuzzy set A is convex if for any λ in $[0,1]$

$$\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \min(\mu_A(x_1), \mu_A(x_2))$$

→ Alternatively A is convex, if all α- cuts are convex



2 Convex Fuzzy set



Non - Convex

FUZZY NUMBERS :

→ a fuzzy no. A is a fuzzy set in $\mathbb{R}$ that satisfies normality & convexity

Bandwidth (BW) :

→ for a normal and convex set the bandwidth is the dist.ⁿ btw. 2 unique crossover pts.

$$BW(A) = |x_2 - x_1| \quad , \text{with} \quad \mu_A(x_1) = \mu_A(x_2) = 0.5$$

Symmetry :

→ fuzzy set A is symmetric if its MF is symmetric around a certain pt. $x = c$,

$$\mu_A(x+c) = \mu_A(x-c) \quad \forall\ x \in X$$

Open left, Open Right, Closed fuzzy set A

OL, OR, CL

$$A_{OL} : \underbrace{\lim_{x \to -\infty} \mu_A(x) = 1}_{K_1} \qquad \underbrace{\lim_{x \to \infty} \mu_A(x) = 0}_{K_2}$$

A_{OR} : $K_1 = 0$ $\qquad K_2 = 1$

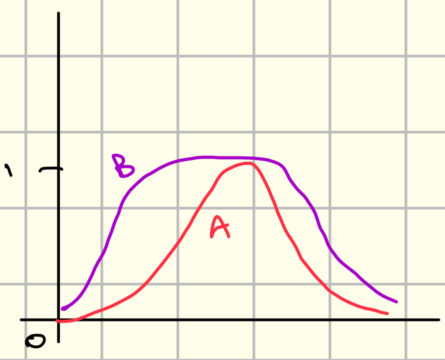
A_{CL} : $K_1 = 0$ $\qquad K_2 = 0$

Subset: $A \subseteq B$: $\mu_A \subseteq \mu_B$

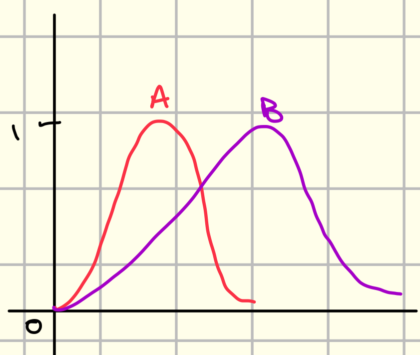
Complement: $\bar{A} = X - A$: $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$

Union : $C = A \cup B$: $\mu_C(x) = \max(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x)$

Intersection : $C = A \cap B$: $\mu_C(x) = \min(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x)$

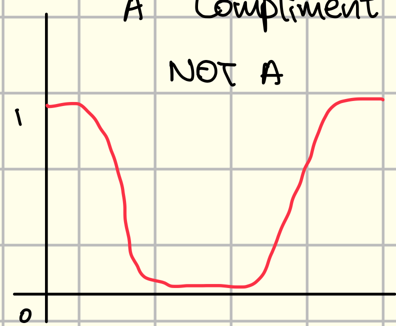


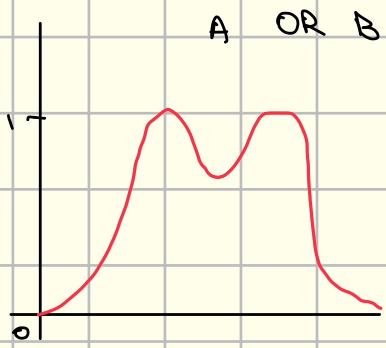
A is contained in B

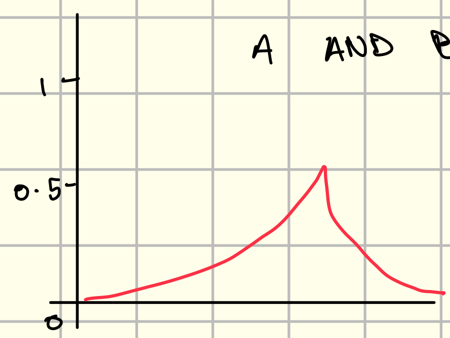


A' Compliment

NOT A







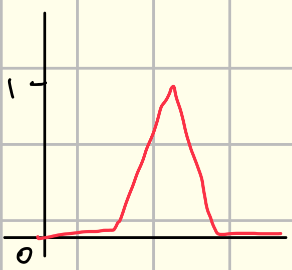
MF in one dimension:

→ Triangular MF: $\text{trimf}(x, a, b, c) = \max\left\{\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right\}$

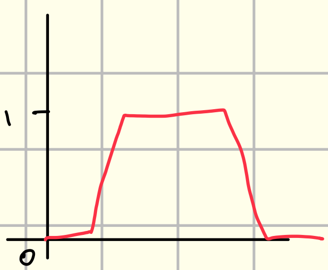
→ Trapezoidal MF: $\text{trapmf}(x, a, b, c, d) = \max\left\{\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right\}$

→ Gaussian MF: $\text{gaussmf}(x, c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2}$

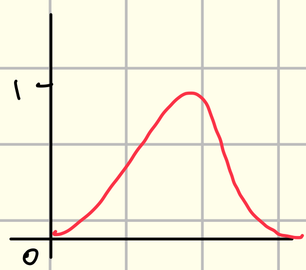
→ Generalized bell MF: $\text{gbellmf}(x, a, b, c) = \dfrac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}}$



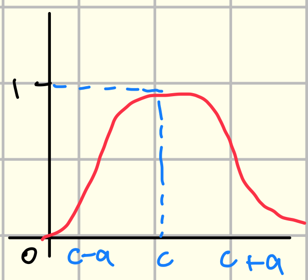
(1)



(2)



(3)



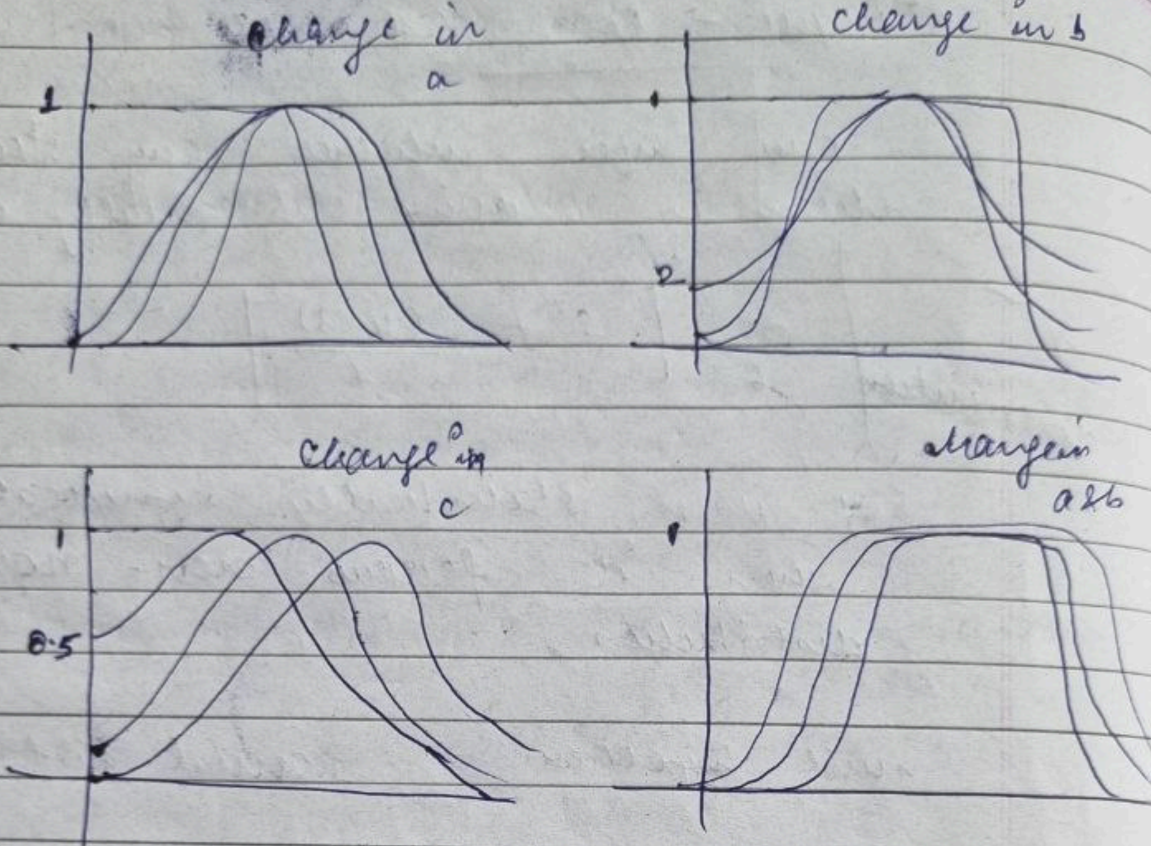
(4)

Artificial Intelligence-II

Generalized bell MF (x, a, b, c)

$$= \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}}$$

change of parameters in bell MF :-

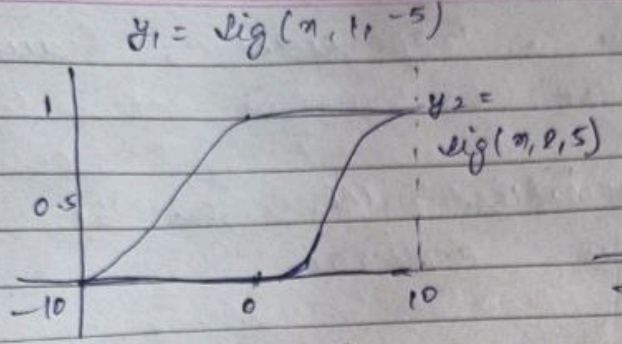


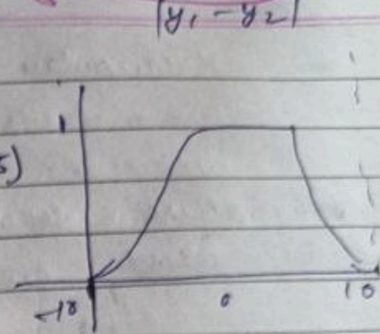
Gaussian MF and Bell MF achieve smoothness, they are unable to specify <u>Asymmetric</u> MFs which are important in many applications

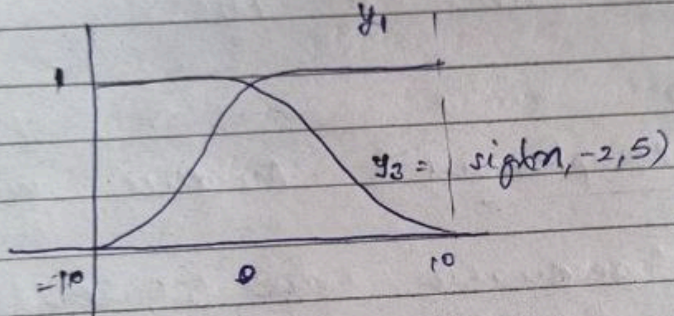
Asymmetric and ~~closed~~ close MFs can be synthesized using either the absolute difference of the product of 2 sigmoid functions.

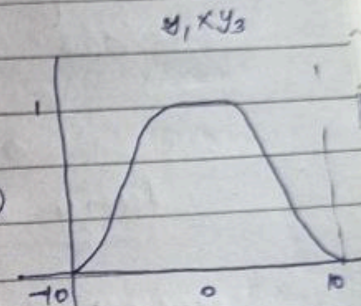
Sigmoidal MFs : $\text{sigmf}(x, a, c) = \dfrac{1}{1 + e^{\left(\frac{a}{x-c}\right)}}$

MF : Matrix factorization









(*) A sigmoid MF is inherently open right or left and thus it is appropriate for representing concepts of very large or very +ve / very -ve.

(*) Sigmoid MF mostly used as activation func^n in ANN

(*) A NN should ~~emphe~~ synthesis a closed MF to simulate the behaviour of a fuzzy sys.

* Triangular, Trapezoidal, Gauss, Bell
  → not exhaustive

* other specialized MFs can be created based on application

* Any type of cont. prob. dist. fun can be used as MFs.

<u>Fuzzy complement</u> : Another way to define reasonable and consistent operation on fuzzy sets.
— General requirement

Boundary : $N(0) = 1$, $N(1) = 0$
monotonicity : $N(a) > N(b)$, if $a < b$
Involution : $N(N(a)) = a$

2 Types of fuzzy complement.

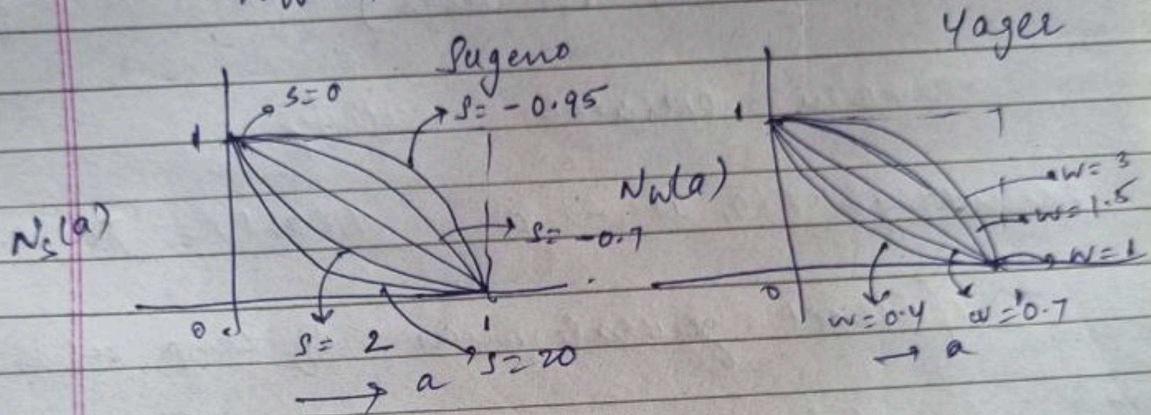
→ SUGENO's complement

$$N_s(a) = \frac{1-a}{1+sa} \quad [s > -1]$$

family of fuzzy complement operator

→ YAGER's complement

$$N_w(a) = (1 - a^w)^{1/w} \quad (w > 0)$$



fuzzy intersection and union, complement

The intersection of 2 fuzzy sets A and B is specified in general by a func^n.

$$T : [0, 1] \times [0, 1] \longrightarrow [0, 1] \text{ with}$$

$$\mu_{A \cap B}(x) = T(\mu_A(x) = \mu_B(x)) = \mu_A(x) \,\tilde{*}\, \mu_B(x)$$

↑ binary operator for func^n T

crisp

This class of fuzzy intersection operators is called T-norm (Triangular operators)

T-norm operator satisfy

Boundary: $T(0, a) = 0$, $T(a, 1) = T(1, a) = a$

(Correct generalization of crisp set)

Monotonicity: $T(a, b) \le T(c, d)$ if $a \le c$ and $b \le d$

D
B

A decrease of membership in A & B cannot increase a membership in $A \cap B$

imp
kuch
samjh
nhi
aa
rha

Commutative $T(a, b) = T(b, a)$

T is indifferent to the order of fuzzy sets to be combined.

Associativity $T(a, T(b, c)) = T(T(a, b), c)$

]

V. II M N

Date NO
Page

Intersection is independent of the order of pair wise grouping.

Example of T- norm

minimum $T_m(a,b) = \min(a,b) = a \wedge b$

Algebraic product : $T_a(a,b) = ab$

Bounded product : $T_b(a,b) = \cancel{0 \vee (a+b)}$
$0 \vee (a+b-1)$

Drastic product : $T_d(a,b) = \begin{cases} a, & \text{if } b=1 \\ b, & \text{if } a=1 \\ 0, & \text{if } a,b<1 \end{cases}$

Imp
kuch
Samjh
nhi
aa
rha

T- conorm or S-norm

$S: [0,1] \times [0,1] \rightarrow [0,1]$

$\mu_{A \cup B} = S(\mu_A(x), \mu_B(x)) = \mu_A(x) \tilde{+} \mu_B(x)$

Boundary : $S(1,1) = 1$, $S(a,0) = S(0,a) = a$

Monotonicity $S(a,b) < S(c,d)$ if $a<c$, $b<d$

c)

Commutativity $s(a,b) = s(b,a)$

Associativity $s(a, s(b,c)) = s(s(a,b), c)$

Examples.

Max : $s_m(a,b) = \max(a,b) = a \vee b$

Algebraic sum : $s_a(a,b) = a + b - ab$

Bounded sum : $s_b(a,b) = 1 \wedge (a+b)$

Drastic sum : $s_d(a,b) = \begin{cases} a, & \text{if } b=0 \\ b, & \text{if } a=0 \\ 1, & \text{if } a,b>0 \end{cases}$

Generalized De morgan's law

$T(a,b) = \cancel{N(S(N(a)))}$
$N(S(N(a), N(b)))$

$S(a,b) = N(T(N(a), N(b)))$

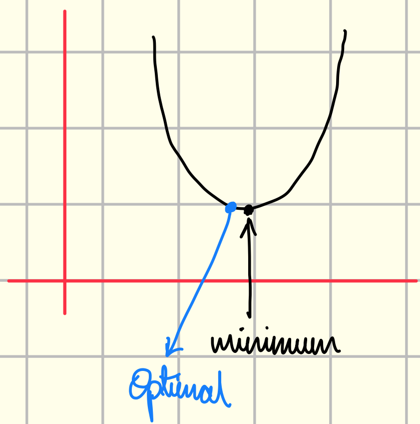
T-norms and S-norms are duals which support the generalization of De morgan's law.

LEARNING PARAMETERS :

→ if our learning parameters are,

$\theta : \{ \theta_0, \theta_1, \ldots\ldots \theta_k \}$

$w : \{ w_0, w_1, \ldots\ldots w_k \}$



$x = ?$ $\{ 3x + 2 = 0$

$$x, y = ? \quad \underbrace{\begin{cases} 3x + 4y = 3 \\ 5x + 3y = 6 \end{cases}}$$

can be solved deterministically

Linear Regression

$SS_{Reg} = S_{\theta_0 \theta_1}$

$\frac{\partial S}{\partial \theta_0} = 0$, $\frac{\partial S}{\partial \theta_1} = 0$

$\hat{\theta}_0 \; \hat{\theta}_1$ → learning Parameters

→ loss = f (actual value, predicted value)

→ we can solve the linear regression deterministically but we can also use Optimization Algorithms

→ Genetic Algorithm comes in between these deterministic & Optimization methods

→ topics covered in AI-1 will also be included in AI-2

→ Revise GA (Genetic Algorithm) from AI-1

GENETIC ALGORITHM (GA):

→ Randomized search & Optimization Technique guided by the principle of Natural genetic System

→ Inspired by biological evolution process

→ uses concepts of 'Natural selection', 'Genetic Inheritence', 'Survival of the fittest' (Darwin 1859)

→ Originally developed by John Holland (1975) & gained popularity in late 80's

WHY GA ?

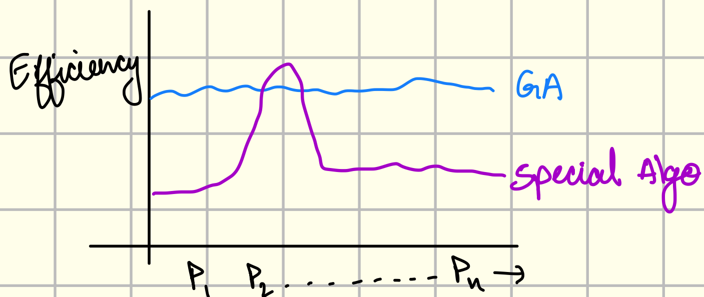
→ Most of the real life problems can't be solved in Poly time using deterministic Algo

→ sometimes near optimal solutions that can be generated quickly are more desirable than optimal sol.n which require huge amount of time.

→ when one problem can be modeled as optimal one

→ Efficient searches for global optima when multiple optima exists.

→ Parallelism is easier



GA Features / Characteristics :

→ Evolutionary Search & Optimization Techinque

→ Principle of Evolution is followed (survival of the fittest & Inheritence)

→ Work with ENCODING of the param. set

→ SEARCHING from a population of points

→ Uses PROB. Transition Rule

GA vs. NATURE :

→ A Solution (Phenotype)

→ Representation of Solution (genotype)

→ Components of Representation

→ Solution Quality (fitness function)

→ Individual (Chromosome)

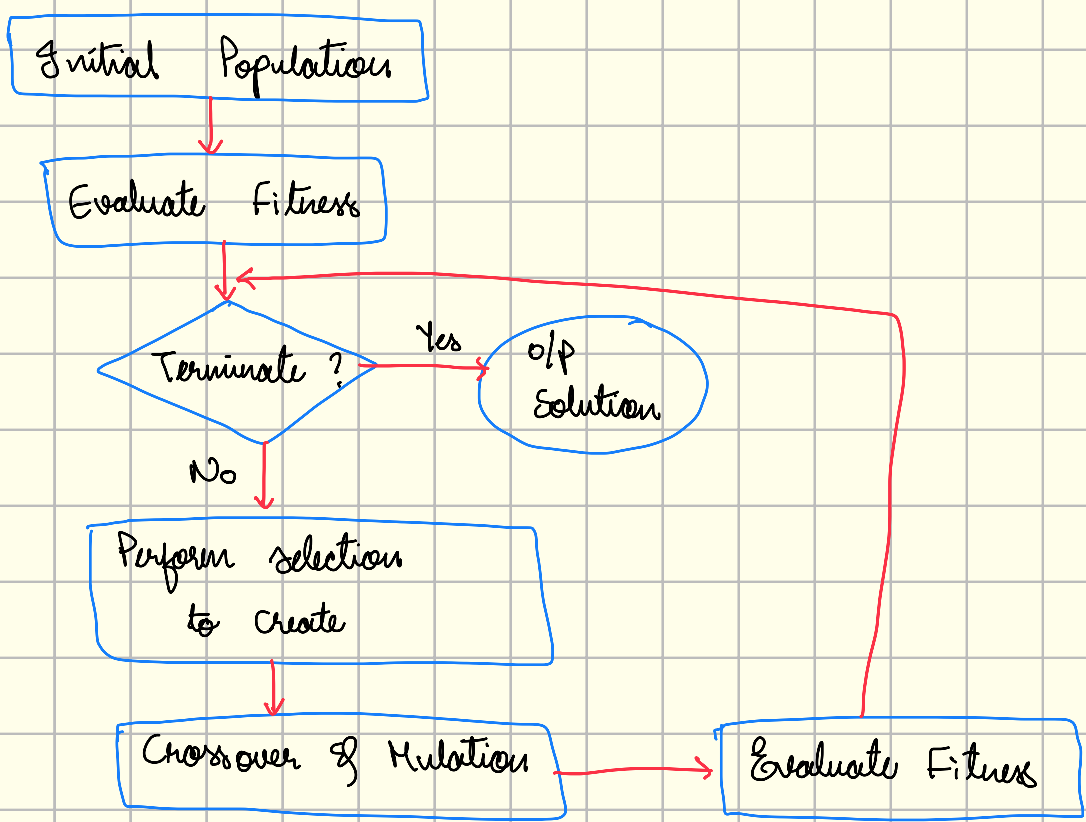
→ Genes (Individuals degree of ability to adapt to surroundings)

→ Selection

→ Crossover (Reproduction)

→ Mutation

GA ALGO :



Simple GA :

→ Produce an initial Population of Individuals

→ Evaluate the fitness of all individuals

→ while Termination criteria not met do

Select fitter individuals for reproduction

Recombine (crossover) btw. individuals

Mutate individuals

Evaluate the fitness of modified individuals

Generate new population

→ End while loop.

Encoding and Population :

→ Chromosomes encodes a solution in the search space

→ Usually as strings of 0's and 1's

→ if l is the string length, no. of different chromosomes (strings) is $\underline{2^{l}}$

→ Population

→ set of chromosomes is a generation

→ Population size is usually constant

→ Common practice is to choose the initial population Randomly.

Assignment-1: Slides

AI Mini Project: Code, Video, Readme File } Github

→ "GAME" with Gaming Agent

ATARI Game

fuzzy

Bayesian

Meta Agent

Deadline: 11^{th} November

Evaluation: $11-15^{th}$ Nov

Assignment-2: Term Paper

Encoding & Population:

Optimize $f(x) = x(8-x) \mid x = [0, 8]$

Binary String of 8-bits

0 - 255 ⟷ 0 - 8

1	0	0	1	1	0	1	0

= 154

$$x = 0 + \frac{8}{255} \times 154 = 4.8313$$

Initial Population
↓
Evaluate fitness
↓
Terminate
↓
Perform Selection to Create mating pool
↓
CrossOver & Mutation
↓
Evaluate fitness

Fitness Evaluation:

→ fitness / Objective funⁿ is associated with each Chromosome

→ this indicates the degree of goodness of the encoded structure

→ the only problem for the specific info (payoff info) that GA uses

→ If the minimization prob. is to be solved then

$$\text{fitness} = \frac{1}{\text{Objective}}$$

→ function : $f(x) = x(8-x)$

Fitness Evaluation :

population : (size = 4)	Corresponding X	fitness / objective fun." : f(x)
10011010	4.8313	15.3089 → f(1)
01100111	3.2313	15.4091 → f(2)
00010101	0.6588	4.8363 → f(3)
10111100	5.8980	12.3975 → f(4)

SELECTION:

→ More copies to good strings

→ fewer " " bad "

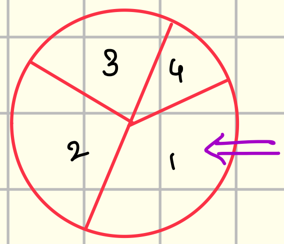
→ proportional selection scheme

→ no. of copies taken to be directly proportional to its fitness

→ Mimics the natural selection procedure to some extent

→ Roulette wheel selection and Tournament selection of frequently used selection procedure.

Chromosome	Fitness
1	15.3089
2	15.4091
3	4.8363
4	12.3975



→ Individual i will have a probability $\frac{f(i)}{\Sigma f(i)}$ to be chosen

Spin 1 :	Chromosome	2	is	selected
S 2 :	"	1	"	"
S 3 →	"	2	"	"
S 4 →	"	4	"	"

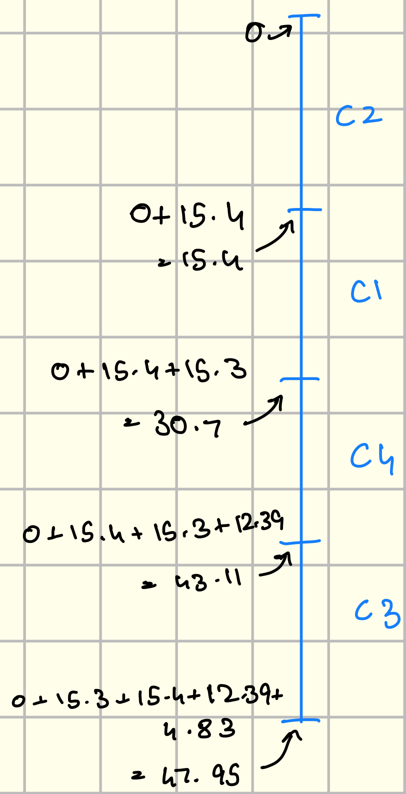
→

Mating Pool

01100111 ← C2
10011010 ← C1
01100111 ← C2
10111100 ← C4

Roulette Wheel Selection:

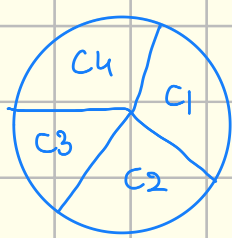
		Fitness
S1 —— C2	C1	15.3
S2 —— C1	C2	15.4
S3 —— C2	C3	4.83
S4 —— C4	C4	12.39



R.W Selection Implementation:

→ Sort the solution in descending order
→ Compute Cumulative sum
→ Generate Random number btw. 0 & 47.95

R.N → 35 → C4
R.N → 2 → C4



Tournament Selection :

1. Repeat until mating pool is full.

→ select a set (size < pop. size) of chromozomes randomly

→ Copy the best chromozome among them into the mating pool

2. USUALLY tournament size is 2 (Binary Tournament)

3. The Chromosome with lowest fitness value can be never copied into the mating pool.

Crossover & Mutation :

→ Exchange of genetic info

→ It takes place between randomly selected parent chromosomes

→ Single point Crossover & Uniform Crossover most common.

→ Probabilistic Operation

Single point Crossover

1	0	0	1	1	0	1	0
0	1	1	1	1	1	0	1

here l (string length) = 8 k (crossover point) = 5

offstring formed

1	0	0	1	1	1	0	1
0	1	1	1	1	0	1	0

Uniform CrossOver

→ Parent

1	0	0	1	1	0	1	0
0	1	1	1	1	1	0	1

Mask: 1 1 0 1 0 1 1 0 (when mask is 1 we change)

Offspring
0 1 0 1 1 1 0 0
1 0 1 1 1 0 1 1

→ Population size: Usually fixed

→ String size: " "

→ Propability of Crossover is μ_c & mutation is μ_m

μ_c is kept high, μ_m is kept low (Think !!)

→ Termination Criteria

→ Hyperparameters are often mutually tuned

→ sometimes adaptive, emperical

→ Hyperparameters

$P = 4$

$l = 8$

realistically

P : $50 \to 100$

μ_c : $[0.6 \to 0.9]$

μ_m : $[0.01, 1]$

l usually depends on req. precision

TERMINATION Criteria

→ The cycle of S, C, M are repeated for no. of times until one of these occur.

→ Average fitness value of population more or less const. over several generations.

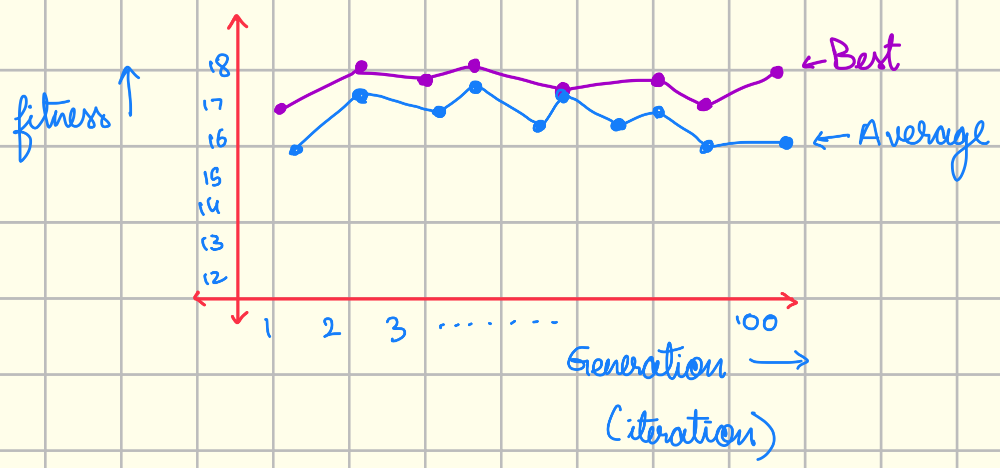
→ Desired obj. fun. value is attained by at least one string in the pop.

→ # Generations (or iterative) is greater than some threshold (commonly used)

ELITIST MODEL:

→ The Best string seen upto the current gen. is preserved in a location either inside or outside the population

Variation of fitness over Generation



MACHINE LEARING:

Prediction	Supervised	Unsupervised	Reinforcement Learning
	{X Y} features, labels $X \in \mathbb{R}^n$ X^i	{X} we cluster the pts. into groups	① Explore knowledge ② Exploit knowledge Mostly used in games Q-Learning & Q-Table

use in Mini Project

SAR

State

Action

Reward

Fully Observable

Partially Observable