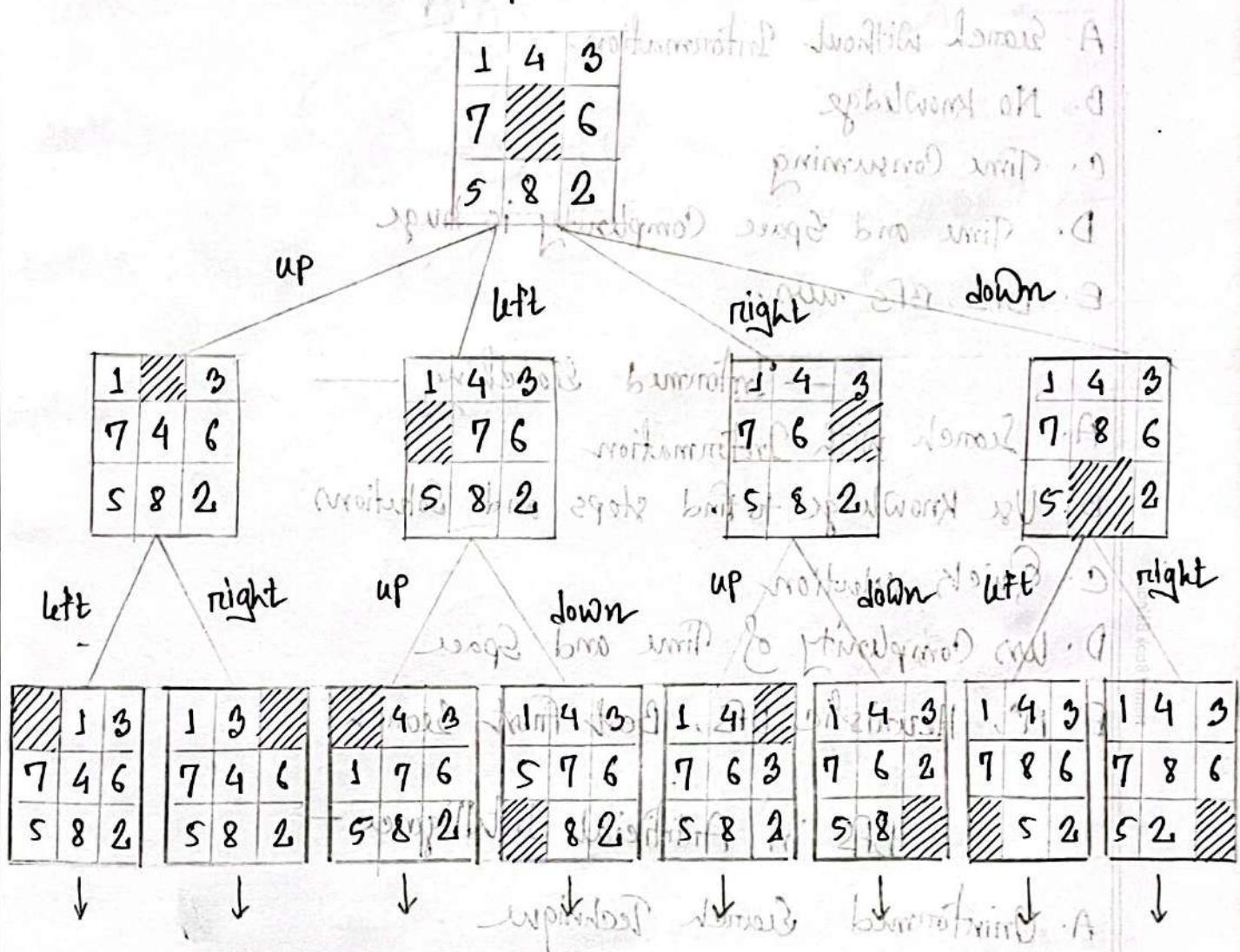


State Space Search



A*

Graph Search \rightarrow Optimal \rightarrow Consistent
 Complete \rightarrow Admissible, Consistent

Tree search \rightarrow optimal \rightarrow Admissible
 (Incomplete) \rightarrow Admissible

Steps = 6

— Uninformed Searching —

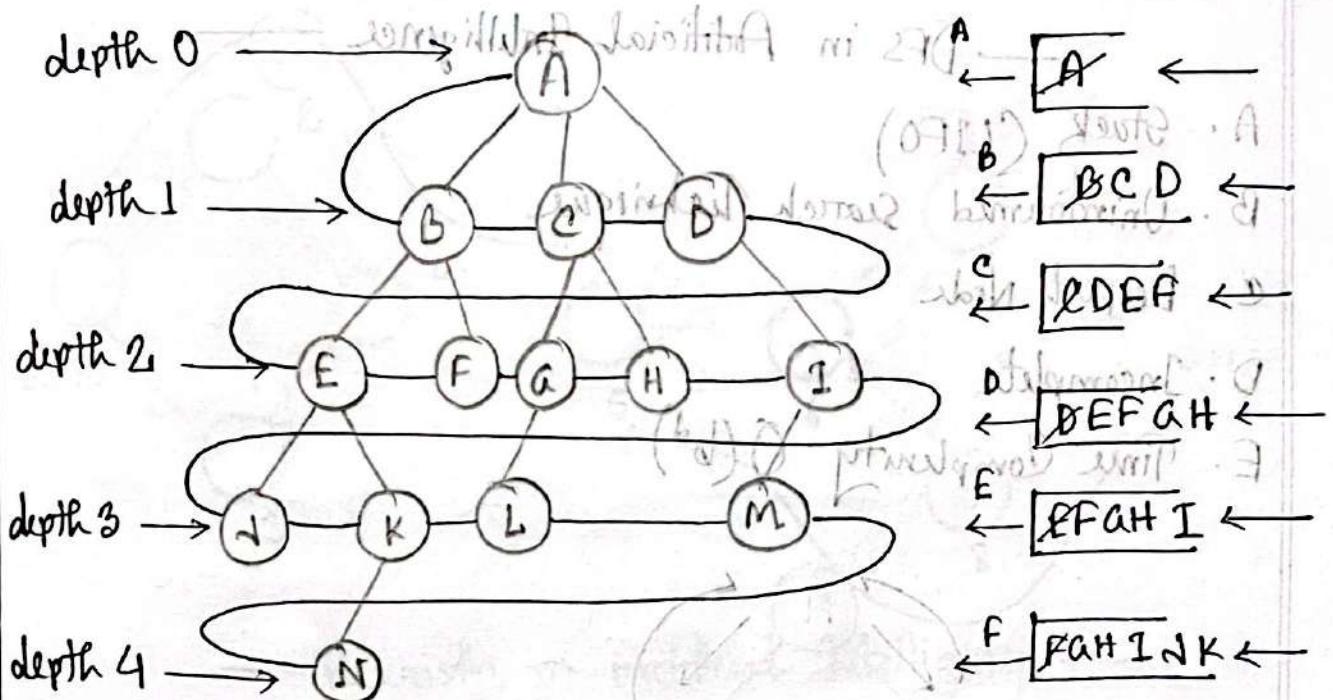
- A. Search without Information
- B. No knowledge
- C. Time Consuming
- D. Time and Space Complexity is huge
- E. DFS, BFS uses

— Informed Searching —

- A. Search with Information
- B. Use knowledge to find steps and solutions
- C. Quick solution
- D. less Complexity of Time and Space
- E. A*, Heuristic DFS, Best first Search

— BFS in Artificial Intelligence —

- A. Uninformed Search Technique
- B. FIFO (Queue)
- C. Shallowest Node
- D. Complete
- E. Time Complexity $O(b^d)$
 - b = Branch factor (child number)
 - d = depth



for child node = 3

$$\text{Complexity} = 3^4 = 81$$

depth 0 \rightarrow 3 A Nodes

depth 1 - (3×3) for BCD Nodes

depth 2 - (3×9) for 27 nodes

depth 3 - (3×15) for 15 Nodes

• Fever shot under a primate. Within a loop must break/lead EMN ME

$$(std) \cdot \frac{M}{MN}$$

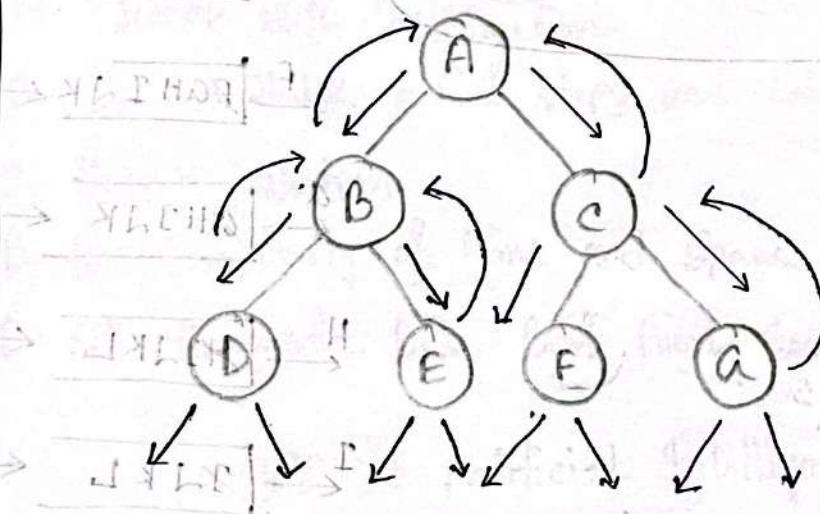
$M \leftarrow \boxed{MN} \leftarrow$

Narrow Joint Allowance Δt

Normal with MyoQuinol

→ DFS in Artificial Intelligence → 0 Steps

- A. Stack (LIFO)
- B. Uninformed Search Technique
- C. Deepest Node
- D. Incomplete
- E. Time Complexity $O(b^d)$



$[AC-GFBED]$ Traversal Nodes Sequence

$b = \text{above limit not}$
 $b_G = \text{max depth}$

$b_8 =$

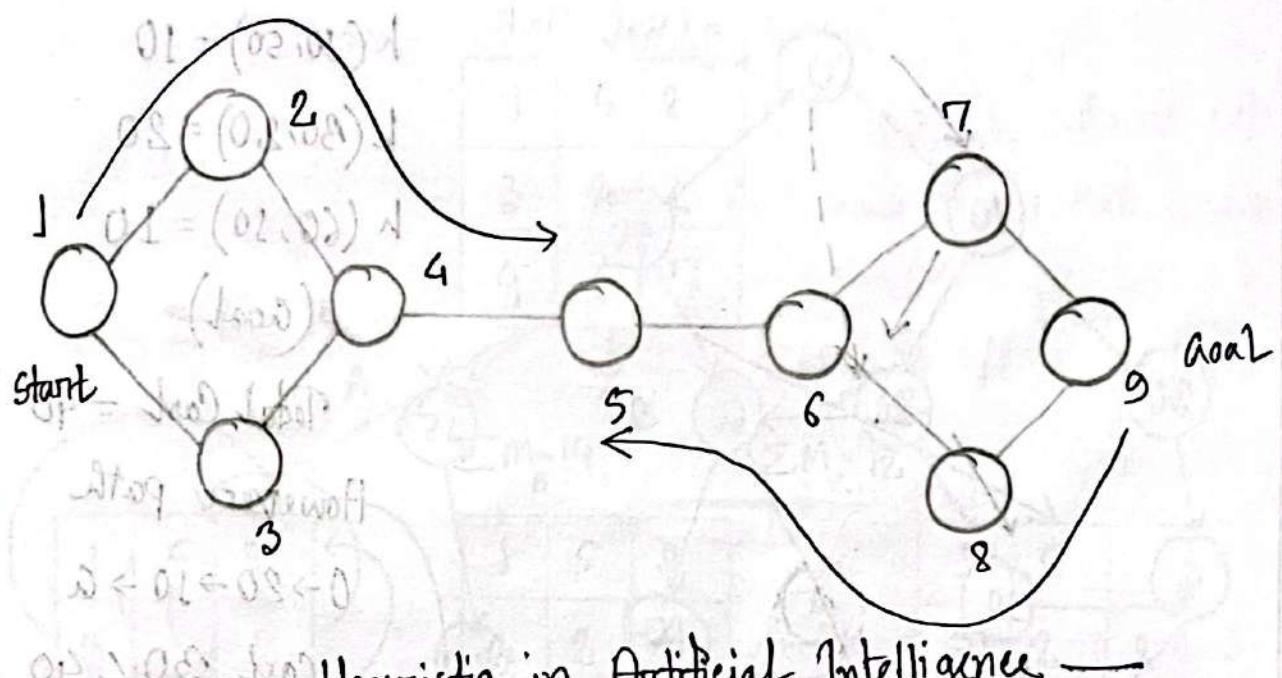
→ Bidirectional Search Algorithm → 0 Steps

Two simultaneous search from an initial node to goal and backward from goal to initial, stopping when two meet.

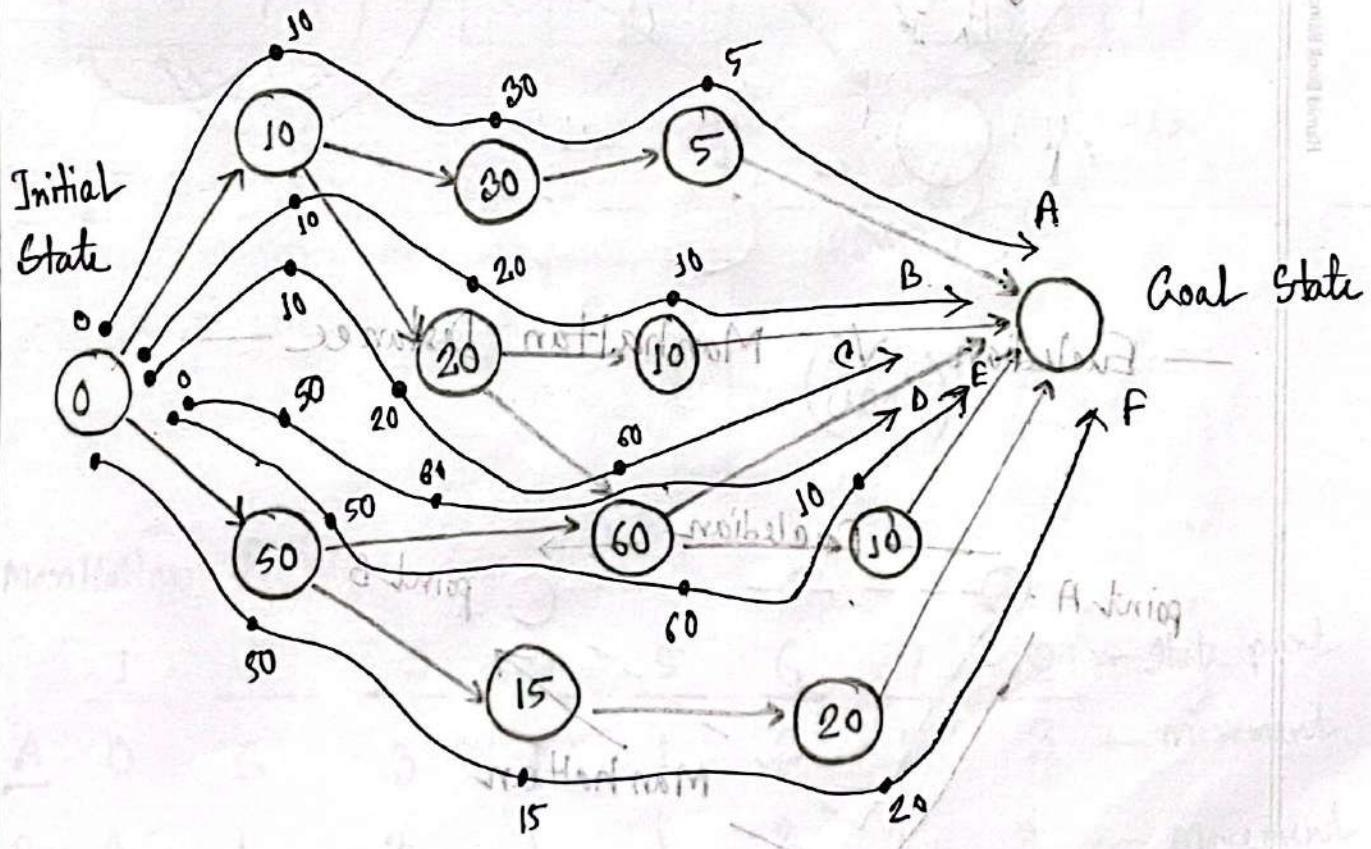
Time Complexity: $2(b^{d/2})$

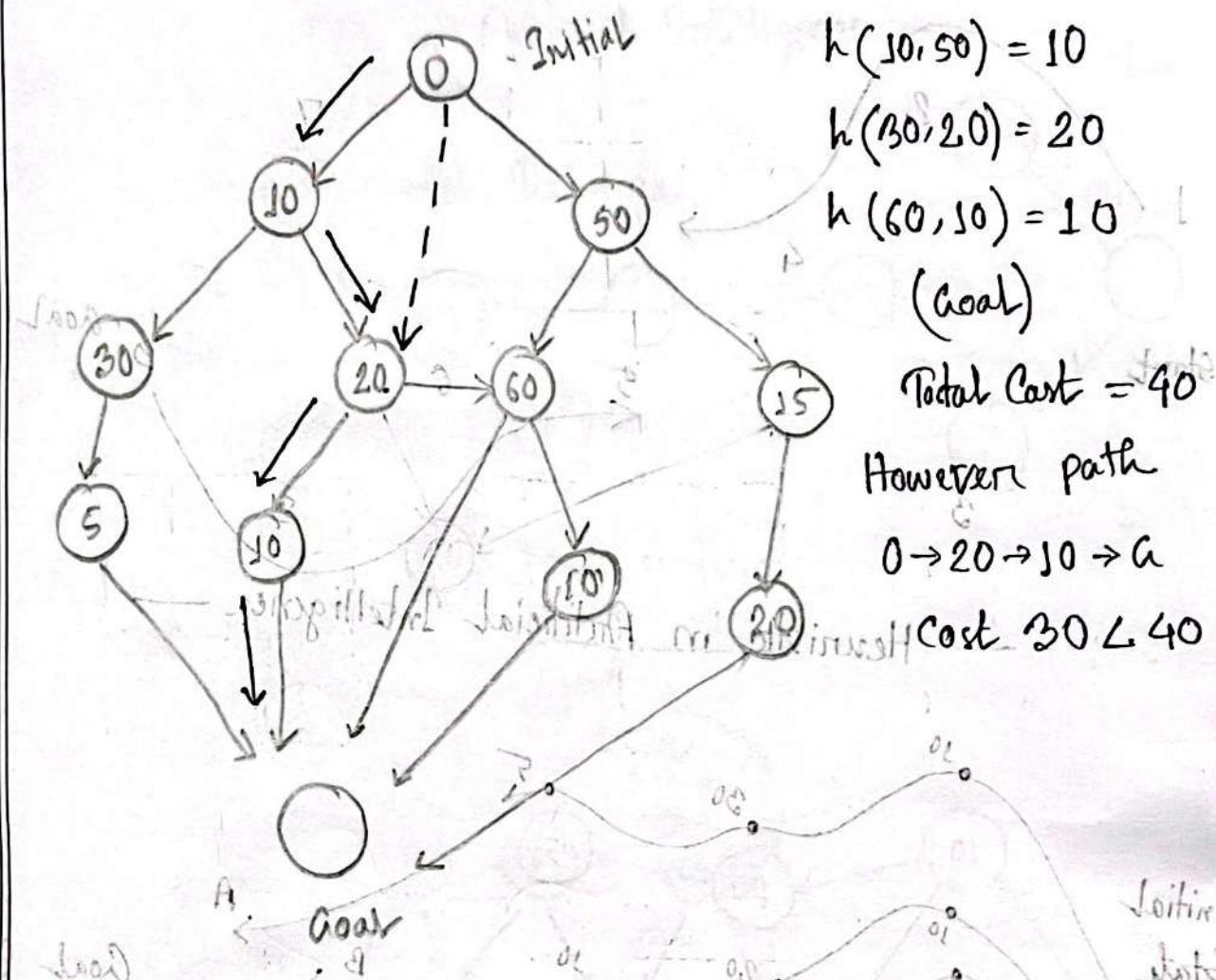
Complete in breadth first search

Not in depth first search

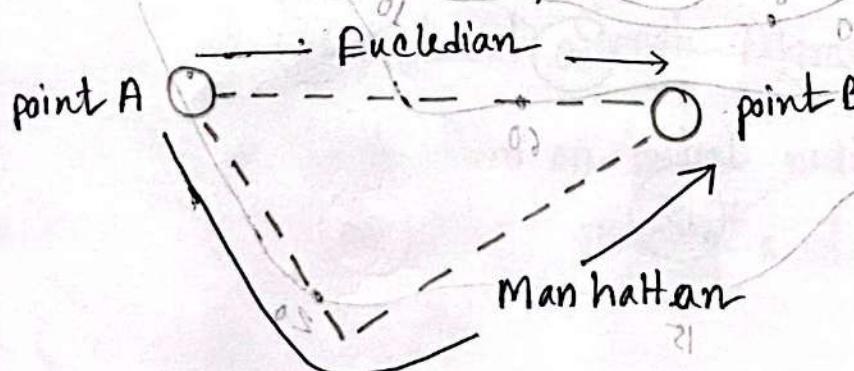


$01 > 02 > 03$ — Heuristic in Artificial Intelligence —





— Euclidean vs Manhattan Distance —



— Nonsol. 3x3 - 8-puzzle —

1	5	8
3	2	
4	6	7

$$\sum M_A = 14$$

A X X B

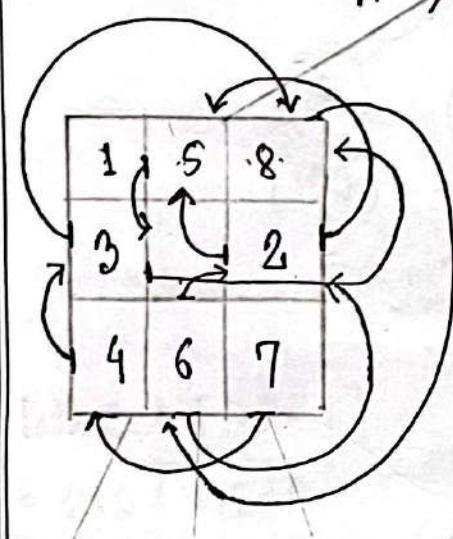
$$\sum M_B = 14$$

$$\sum M_C = 12$$

1	5	8
3	2	
4	6	7

1	5	8
3	2	7
4	6	

1	5	8
3	2	7
4	6	7



HS

PC

811

101

111

Not 1	Not 2
811	111
001	001

1	2	3
4	5	6
7	8	

Not 1

Not 2

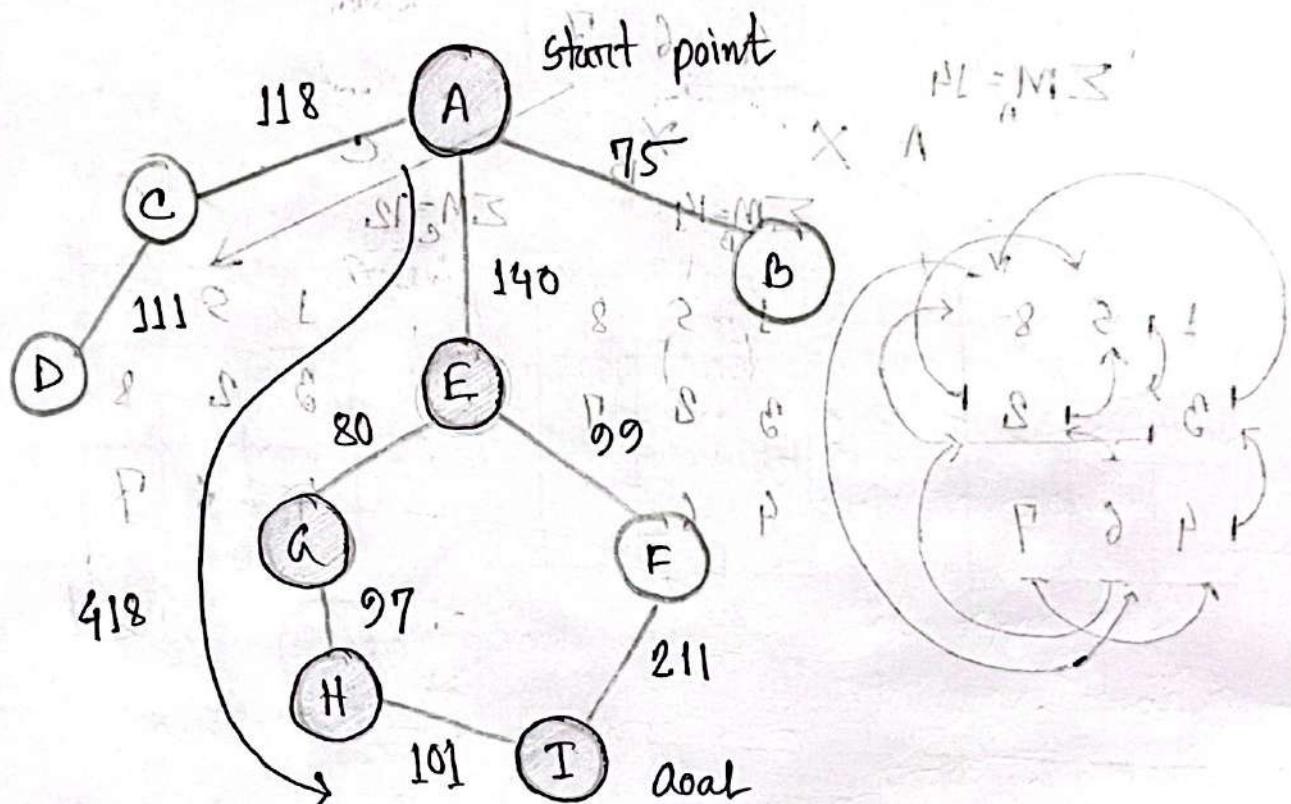
Manhattan Priority Function

	1	2	3	4	5	6	7	8	state point
A	0	2	3	1	5	2	2	3	movement
B	0	1	3	1	1	2	3	3	movement
C	0	1	3	1	1	2	2	2	movement

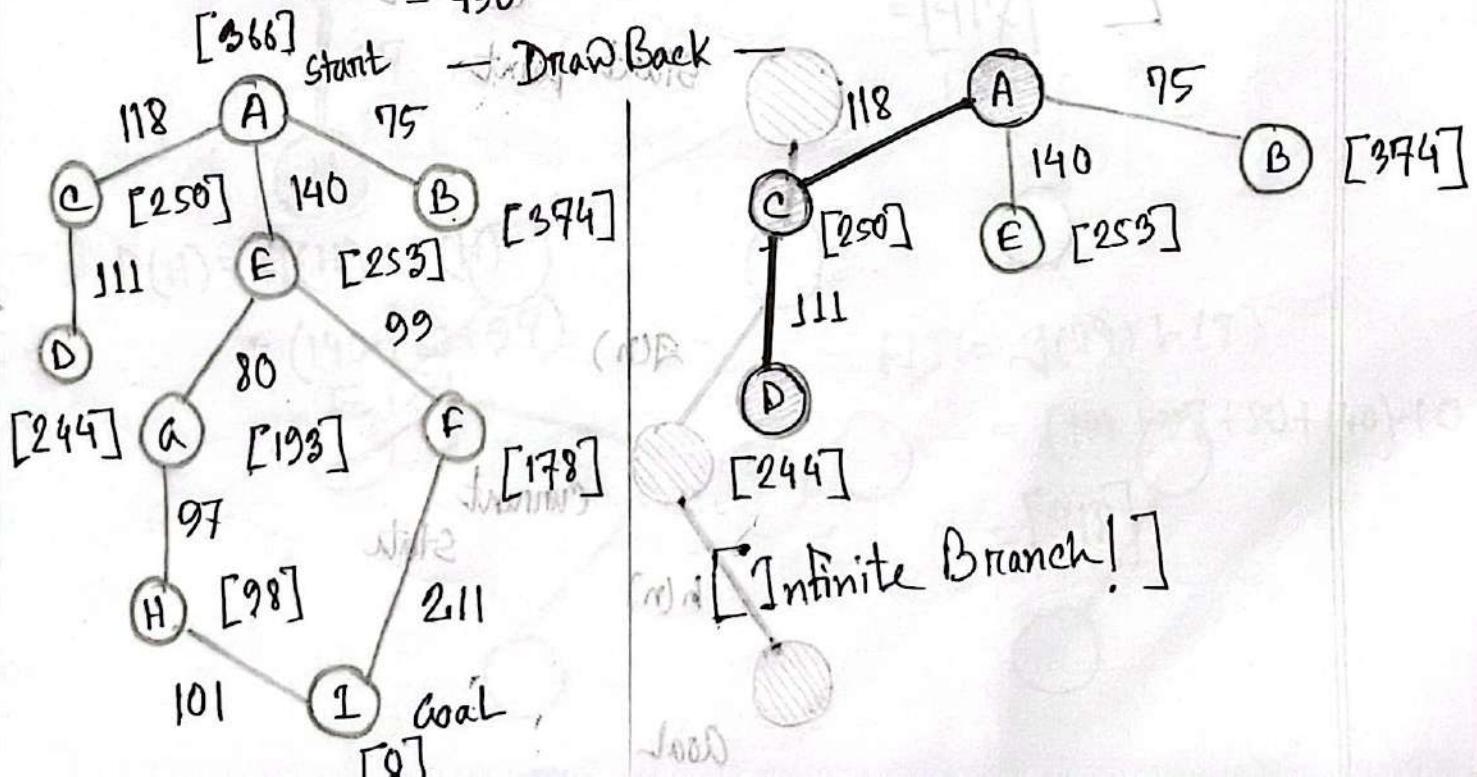
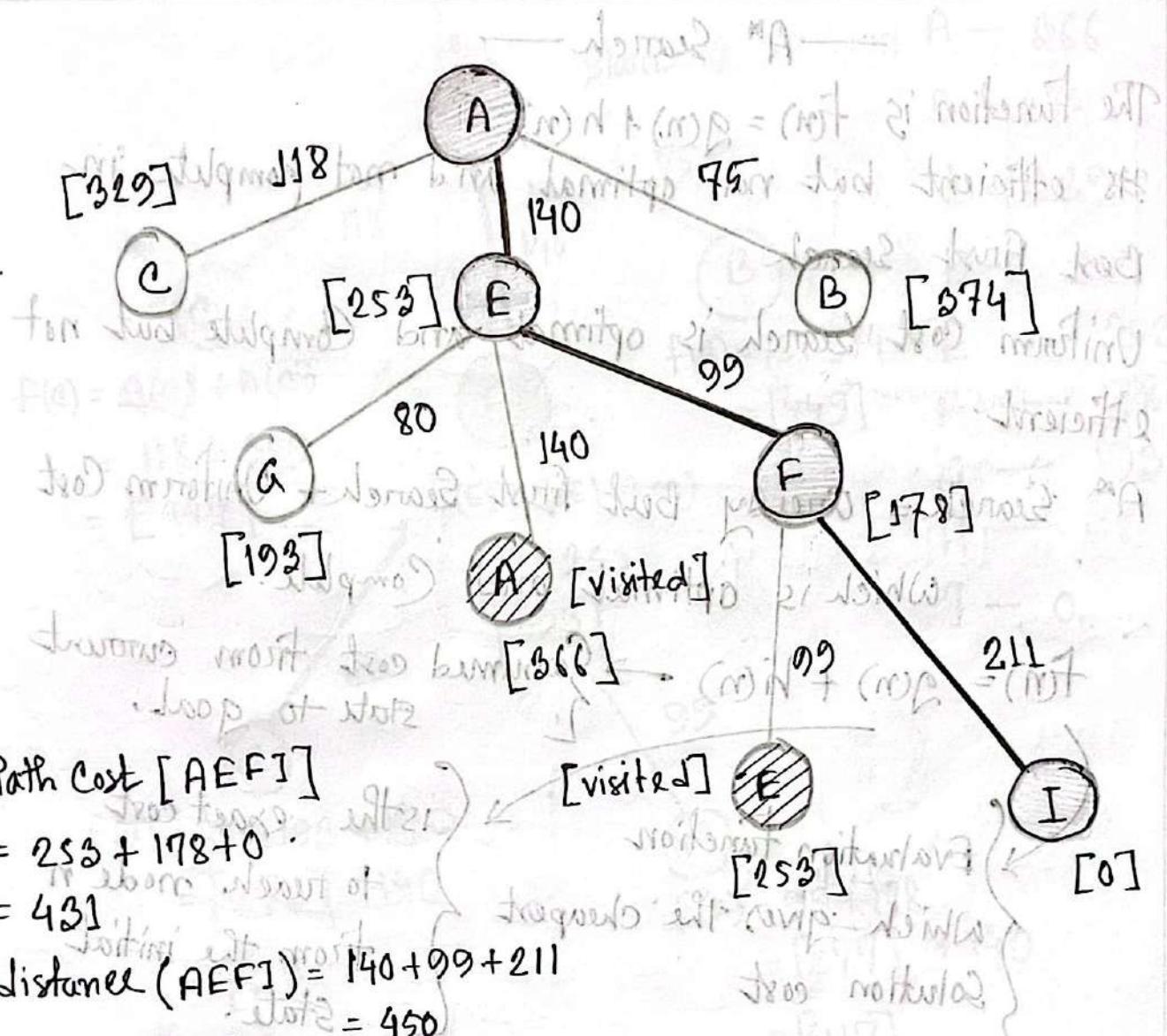
— Greedy Best-First Search —

The function is $f(n) = h(n)$

Where $h(n)$ = estimated cost from node n to the goal.



<u>State</u>	<u>Heuristic h(n)</u>	<u>State</u>	<u>Heuristic h(n)</u>
A →	386	F →	178
B →	374	G →	193
C →	329 ✓	H →	98 Hellmann
D →	244	I →	0
E →	253 ✓	G →	0
F →	0	I →	0
G →	0	G →	0
H →	0	I →	0



— A* Search —

The function is $f(n) = g(n) + h(n)$

It's efficient but not optimal and not complete [incomplete]

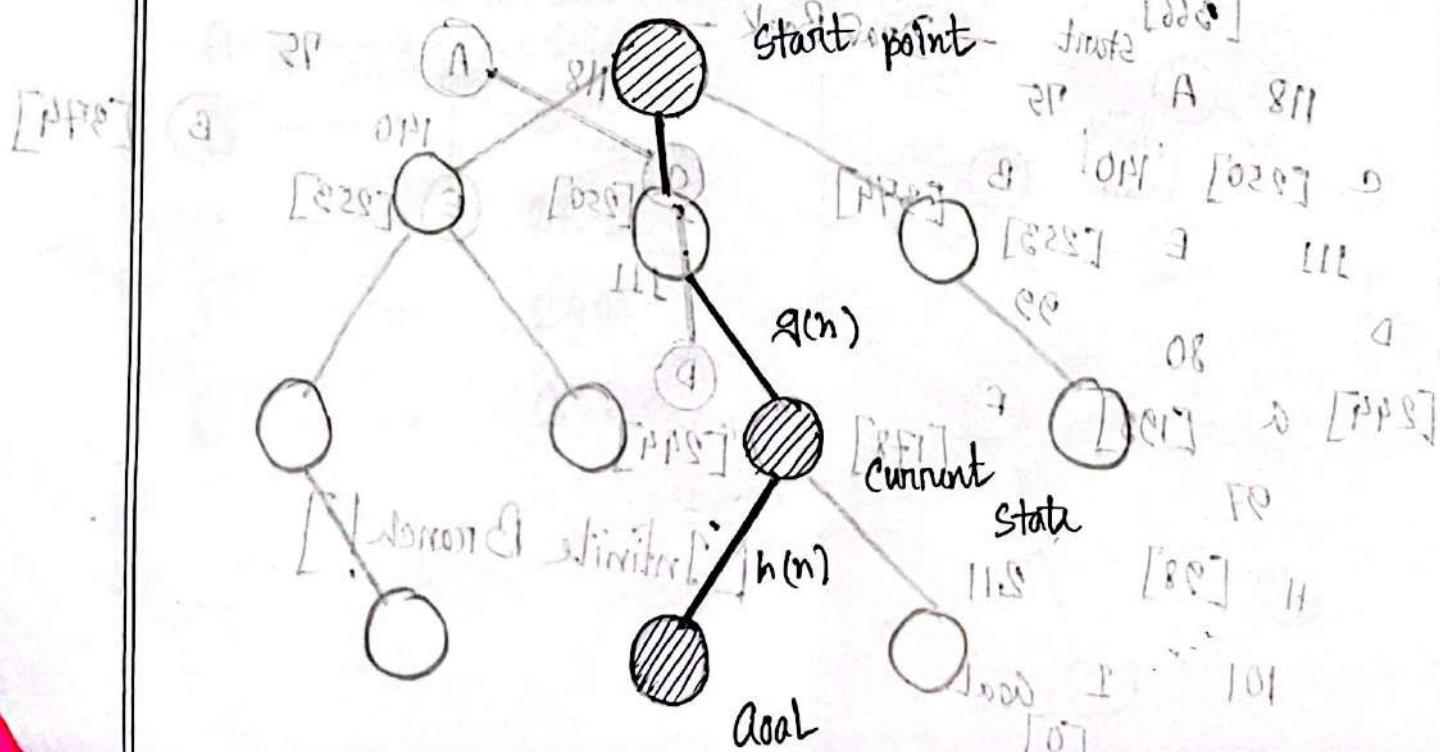
Best first search

Uniform Cost Search is optimal and complete but not efficient.

~~A* Search = Greedy Best first Search + Uniform Cost~~
which is optimal and complete.

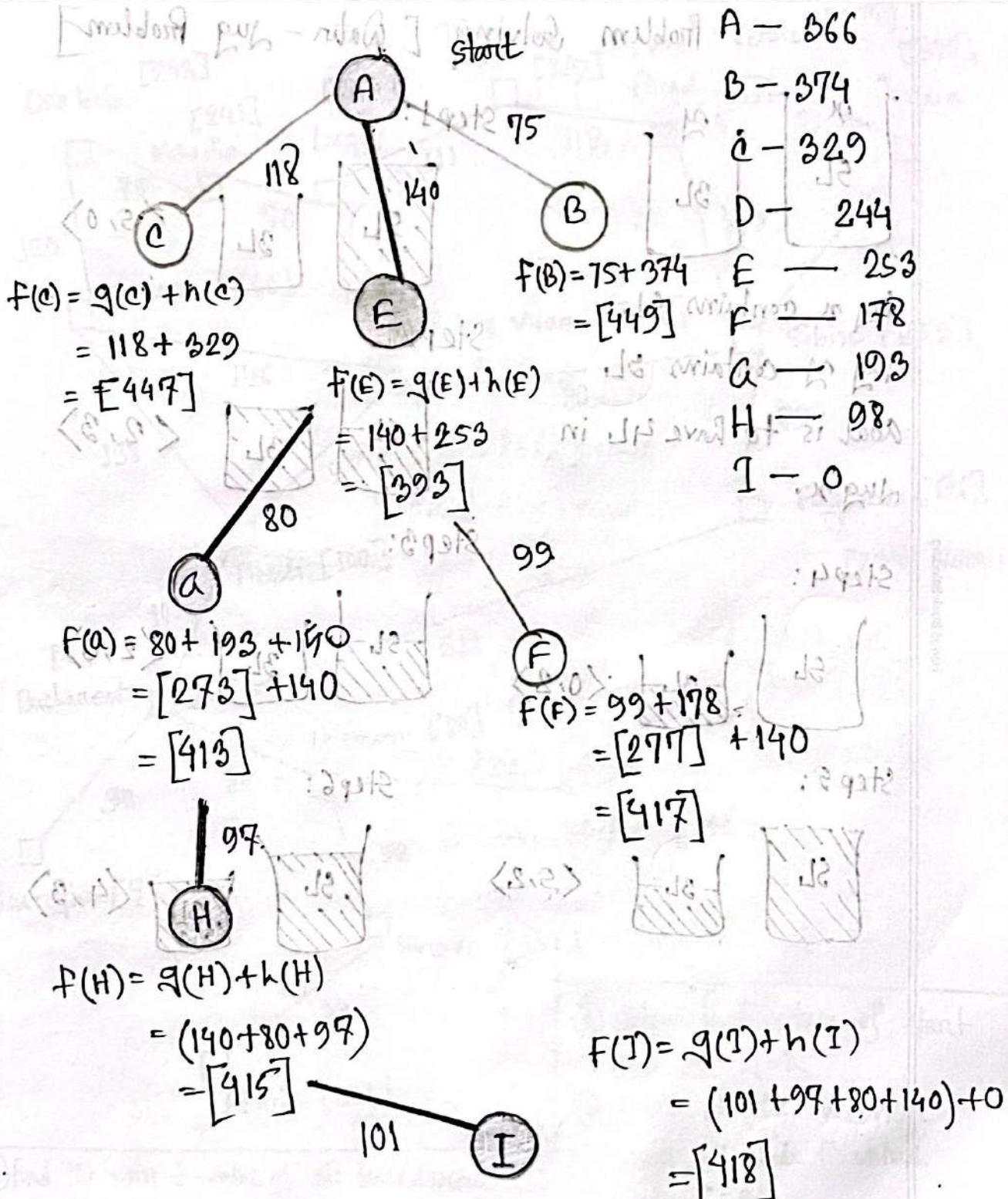
$f(n) = g(n) + h(n)$ → Summed cost from current state to goal.

} Evaluation function
 } which gives the cheapest solution cost
 } is the exact cost to reach node n from the initial state.

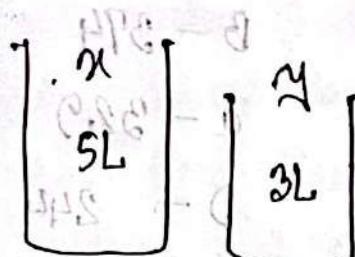


Subject :

Date :



HPS - A Problem Solving [Water - Jug Problem]

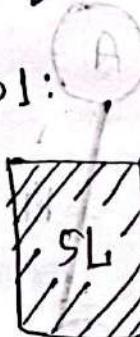


Jug x contains 5L

Jug y contains 3L

Goal is to have 4L in Jug x.

Step 1:



<5, 0>

Step 2:



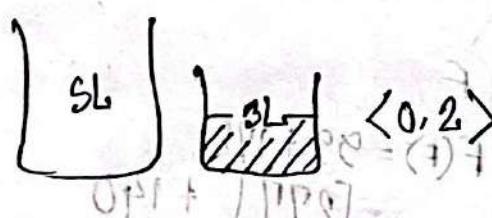
<2, 3>

Step 3:



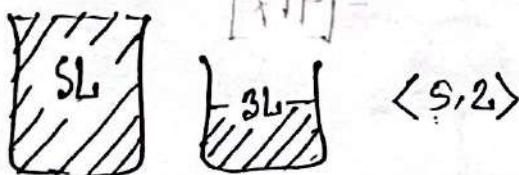
<2, 3>

Step 4:



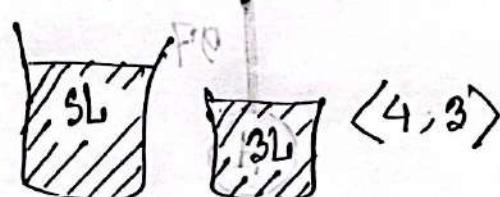
<2, 0>

Step 5:

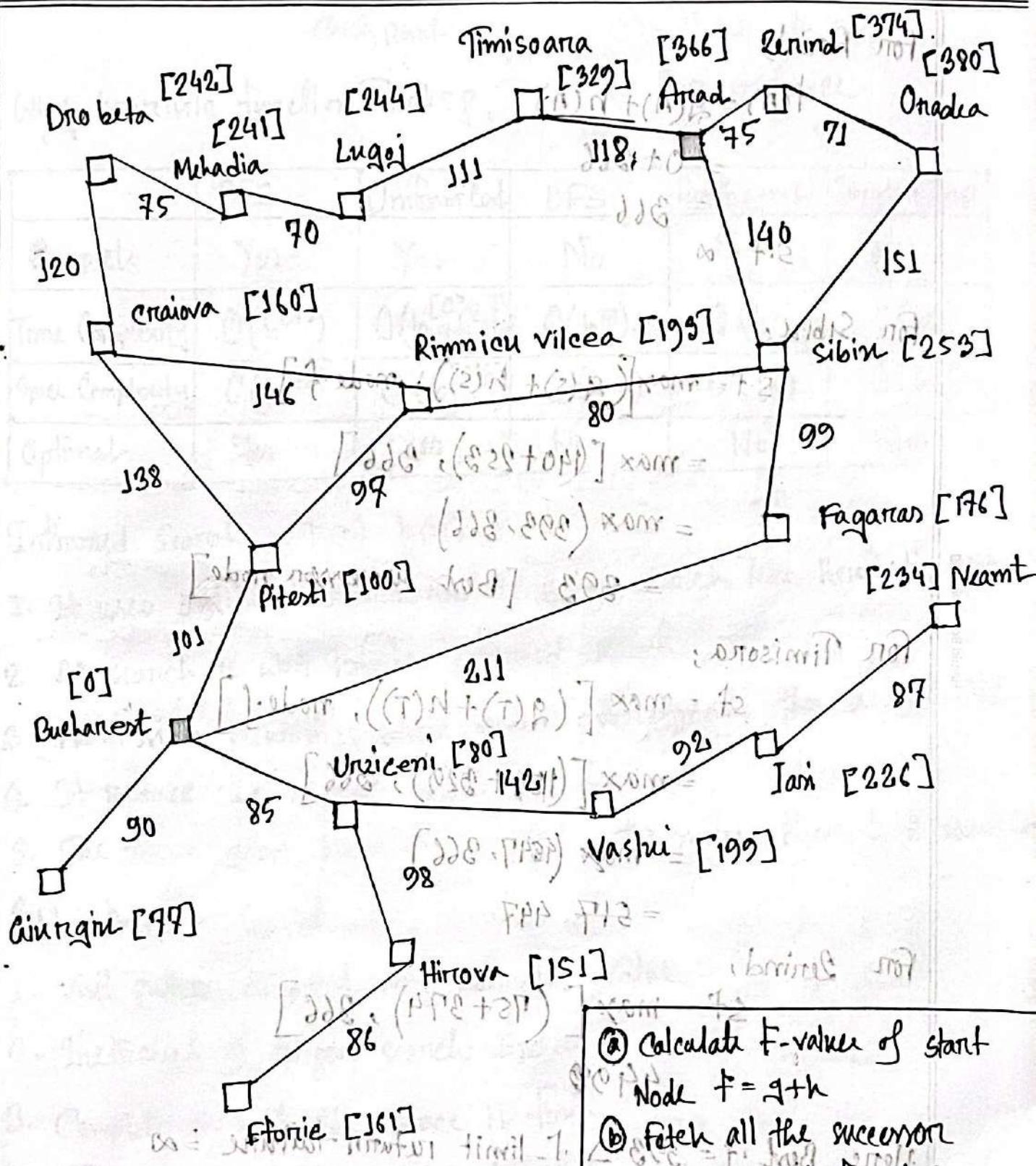


<5, 2>

Step 6:



<4, 3>



- ④ find the min f-value of all successors
⑤

② Calculate F-value of start Node $F = g + h$

⑥ fetch all the successors
and calculate f value
 $f = g + h$

② find $\max(f_{\text{step}}, f_{\text{parent}})$

for Arad;

$$f(A) = g(A) + h(A)$$

$$= 0 + 366$$

$$= 366$$

$$s.f = \infty$$

for Sibiu;

$$s.f = \max [g(s) + h(s), \text{node.f}]$$

$$= \max [(40 + 253), 366]$$

$$= \max (293, 366)$$

$$= 393 \quad [\text{Best successor node}]$$

for Timisora;

$$s.f = \max [g(T) + h(T), \text{node.f}]$$

$$= \max [(189 + 320), 366]$$

$$= \max (489, 366)$$

$$= 489$$

for Zerind;

$$s.f = \max [g(Z) + h(Z), 366]$$

$$= 75 + 374$$

Here Best f = 393 \Rightarrow f-limit returning failure = ∞

$$f_{\text{alt}} = 1$$

$$\text{Alt node} = 449, 447$$

$$f_{\text{limit}} = \min (f_{\text{f-limit}}, \text{alt.f}) = \min (\infty, 449) = 449, 447$$

Check point

 $C^* = \text{Total path cost}$

Why heuristic function works?

 $\ell = \text{Avg step size.}$

	BFS	Uniform Cost	DFS	Depth Limit	Iterative Dup
Complete	Yes	Yes	No	No	Yes
Time Complexity	$O(b^{d+1})$	$O(b^{[c/\ell]})$	$O(b^m)$	$O(b^d)$	$O(bd)$
Space Complexity	$O(b^{d+1})$	$O(b^{[c/\ell]})$	$O(bm)$	$O(bl)$	$O(bd)$
Optimal	Yes	Yes	No	No	Yes

Informed Search : $f(n) = h(n)$

1. It uses additional information to guide search like heuristics value
2. A* search is well known informed search
3. Admissible heuristics value never overestimate the actual cost
4. It reduce the search space
5. The more good heuristics value, the more efficient it would be

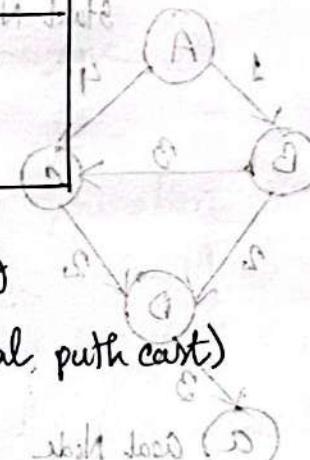
6. Uninformed Search \Rightarrow number of nodes \rightarrow Inefficiency

1. visit paths without any heuristics value
2. Inefficiency of longer search space
3. Complete only if the space is finite
4. find shortest path in BFS if the graph is unweighted
5. Explore the entire search space with worst case
6. Both Tree and graph search

1. BFS is good for shortest path
2. DFS is good for deep search
3. A* for efficiency.
4. Due to heuristics value informed search uses less states than that of uninformed search.
5. IDS (Iterative deepening Search) uses DFS for low memory uses and BFS for completeness simultaneously.
6. A poor heuristics value can mislead the search and a good heuristics can lead the A* search.
7. Overestimating the cost, may cause A* to skip the optimal path. So, heuristics never overestimate the actual path cost.
8. Bidirectional Search to reduce the search space as it can search forward from the start and backward from the goal.

Sudoku Puzzle

$(n)^N \leq (m)$	$(n)^N \leq (m)$	$(n)^N \leq (m)$
$(n)^N \leq (m)$	$(n)^N \leq (m)$	$(n)^N \leq (m)$
$(n)^N \leq (m)$	$(n)^N \leq (m)$	$(n)^N \leq (m)$



Consistency Check $h(n) \leq c(n,m) + h(m)$

$h(A) = 3, h(B) = 2, c(A,B) = 1$ (Actual path cost)

$3 \leq 1 + 2$ True

$h(B) = 2, h(C) = 2, c(B,C) = 3$

$2 \leq 3 + 2$ True

$h(B) = 2, h(D) = 1, c(B,D) = 2$

$2 \leq 2 + 1$ True

$h(C) = 2, h(D) = 1, c(C,D) = 2$

$2 \leq 2 + 1$ True

$h(D) = 1, h(A) = 0, c(D,A) = 3$

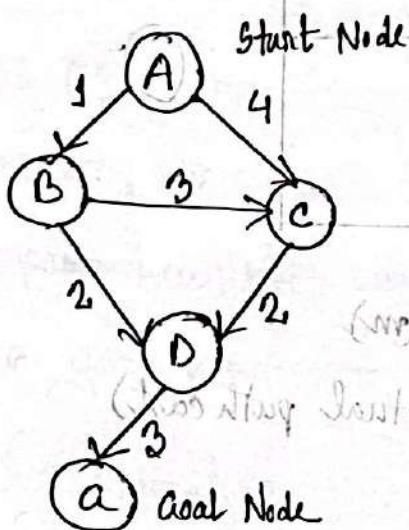
$1 \leq 3 + 0$ True

Consistency satisfied.

Admissible Graph

Conditions:

- (a) Non-negative: $h(n)$ is admissible if $h(n) \geq 0$ for all nodes
- (b) Never overestimate the actual cost: $h(n) \leq h^*(n)$

Underestimate $h(n) \leq h^*(n)$ | Here $h(n)$ = estimated valueOverestimate $h(n) \geq h^*(n)$ | $h^*(n)$ = actual valueHeuristic Table ($h(n)$)

equation: Minimum edge x steps to G

 $A \rightarrow 3$ steps and 1 edge = 3 $B \rightarrow 2$ steps and 1 edge = 2 $C \rightarrow 2$ steps and 1 edge = 2 $D \rightarrow 1$ step and 1 edge = 1 $G \rightarrow 0$ step and 0 edge = 0Actual Cost ($h^*(n)$) $A \rightarrow C \rightarrow D \rightarrow G = 9$ $B \rightarrow D \rightarrow G = 5$ $C \rightarrow D \rightarrow G = 5$ $D \rightarrow G = 3$ $G = 0$

Admissible check

$h(n) = 3, h^*(n) = 9$

$h(n) \leq h^*(n) \checkmark$

$h(n) = 2, h^*(n) = 5$

$h(n) \leq h^*(n) \checkmark$

$h(n) = 2, h^*(n) = 5$

$h(n) \leq h^*(n) \checkmark$

$h(n) = 1, h^*(n) = 3$

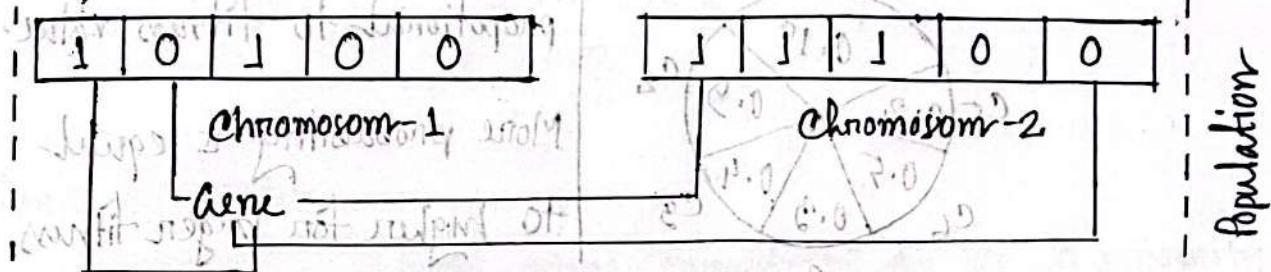
$h(n) \leq h^*(n) \checkmark$

$h(n) = 0, h^*(n) = 0$

$h(n) \leq h^*(n) \checkmark$

So, the graph is admissible

Genetic Algorithm



A. Number of genes per chromosome

B. The codes value

C. The size of population per generation

D. The Crossing over probabilities

E. Mutation probabilities

F. Termination Criteria

G. Selection of Chromosome

A. Roulette Wheel Selection

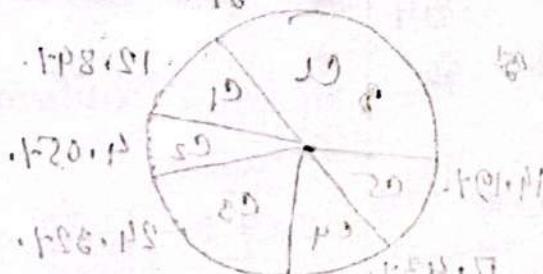
B. Rank Selection

C. Tournament Selection

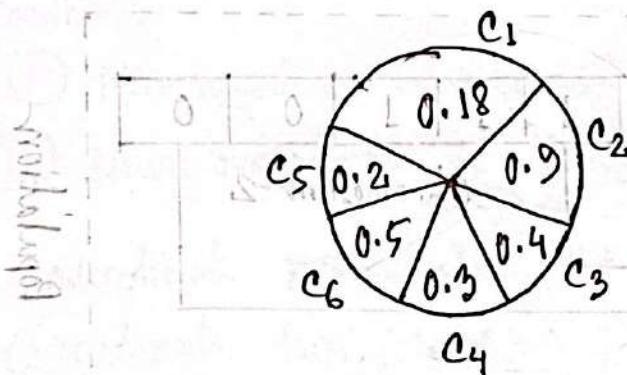
$$P_{SEL} = \frac{C_i}{SPT} = \frac{10}{40} = \frac{1}{4}$$

$$P_{SEL} = \frac{1}{4}$$

$$P_{SEL} = \frac{1}{4}$$



Roulette Wheel Selection:



length of circumference is proportional to fitness value.

More probability is equal to higher for larger fitness.

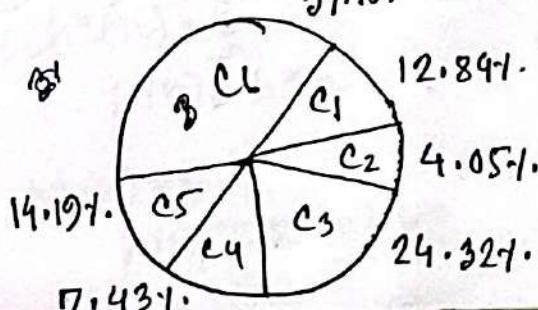
- Determine percentage of Roulette Wheel and Actual Count for six chromosomes with fitness value of 19, 6, 36, 11, 21 and 55.

Chromosome	Fitness	% of wheel	Probability	Expect Count	Actual Count
C ₁	19	0.1284	12.84%	0.77	1
C ₂	6	0.0405	4.05%	0.243	0
C ₃	36	0.2432	24.32%	1.46	2
C ₄	11	0.0743	7.43%	0.44	1
C ₅	21	0.1419	14.19%	0.85	1
C ₆	55	0.3716	37.16%	2.23	2
$\sum F = 148$					

$$\text{Probability Count}_i = \frac{F_i}{\sum F} \quad \text{For } C_1 = \frac{19}{148} = 0.1284$$

$$= 12.84\%$$

Pie chart:



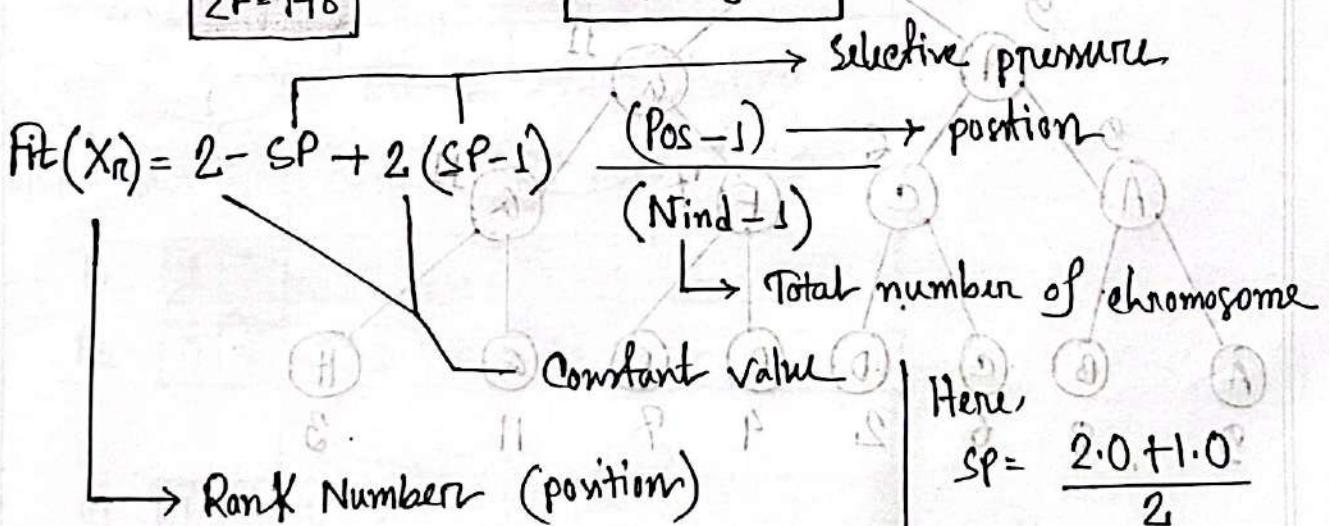
Expected Count = Probabilities \times Total Chromosome

$$\text{for, } c_1 = 0.1284 \times 6 \\ = 0.77$$

Rank Selection:

The less fitness values considered as the priority rank from the previous table; Selective Pressure [1.0, 2.0]

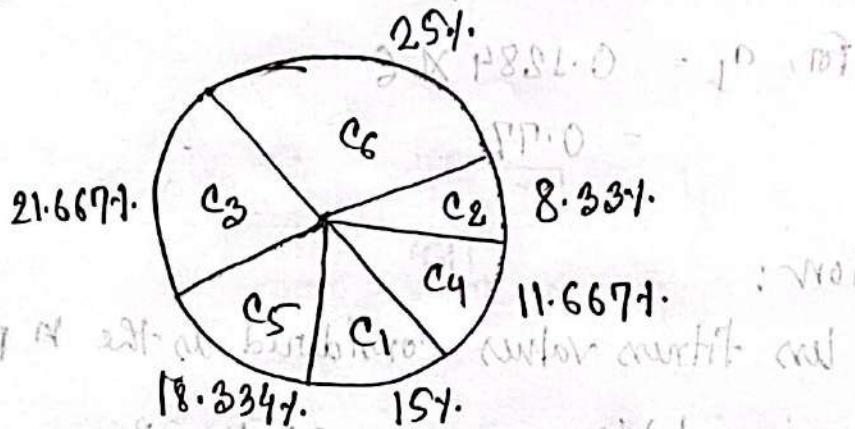
Chromosome	Fitness	Rank	New fitness	Previous %	New %
C ₁	10	3	0.9	12.84%	15.1%
C ₂	6	1	0.5	4.05%	8.33%
C ₃	36	5	1.3	24.32%	21.667%
C ₄	11	2	0.7	7.43%	11.667%
C ₅	21	4	1.1	14.19%	18.334%
C ₆	55	6	1.5	37.16%	25.1%
$\Sigma F = 148$		$\Sigma nF = 6$			



Selective pressure = [1.0, Nind - 1]

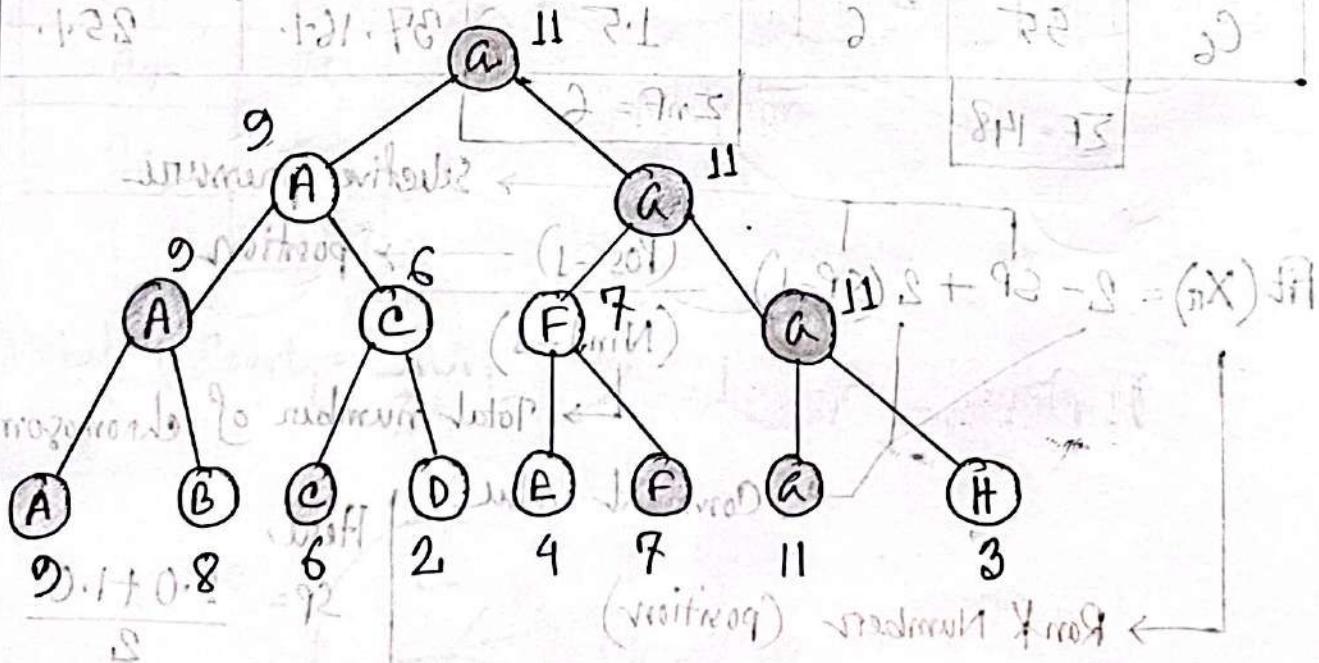
$$= 1.5$$

Pie chart: ~~percentage of Just for self-entertainment = 30%~~ before self



Tournament Selection:

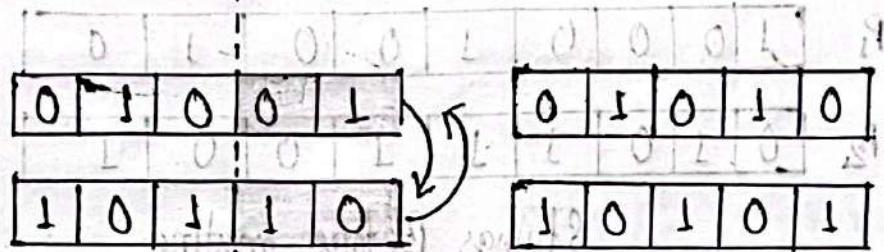
- ① Select k individuals from the population and performs a tournament among them.
 - ② Select the best individual from the k individuals.
 - ③ Repeat the process 1 and 2 until we have the desired amount of population.



$[1 - \ln(1 - 0.1)]$ = winning probability

Crossover And Mutation

Crossover

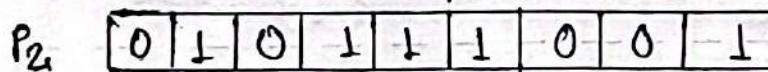
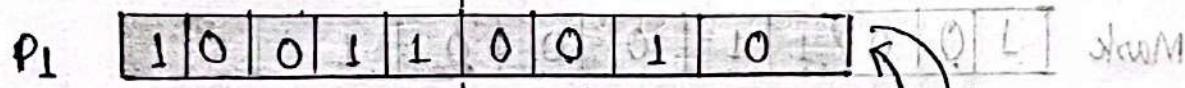


Ⓐ Single-Site Crossover

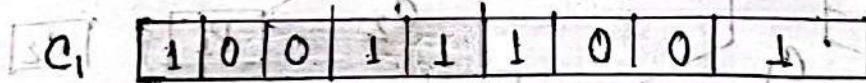
③ Two-site Crossover

④ Crossover Mask

Single-Site Crossover: (At point 5)

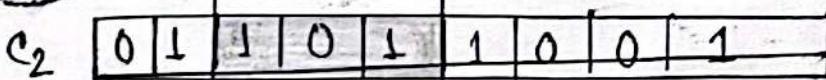
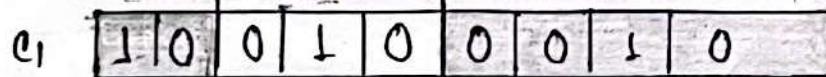
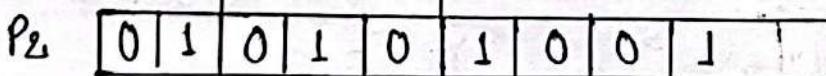


Strings Before mating



Strings: After mating

Two-site Crossover: (At point 2 and 5)



Crossover Mask ∇ noiseless M brief convolution

P₁ 100010010

P_2 0 1 0 1 1 1 0 0 1

15010 Strings Before mating

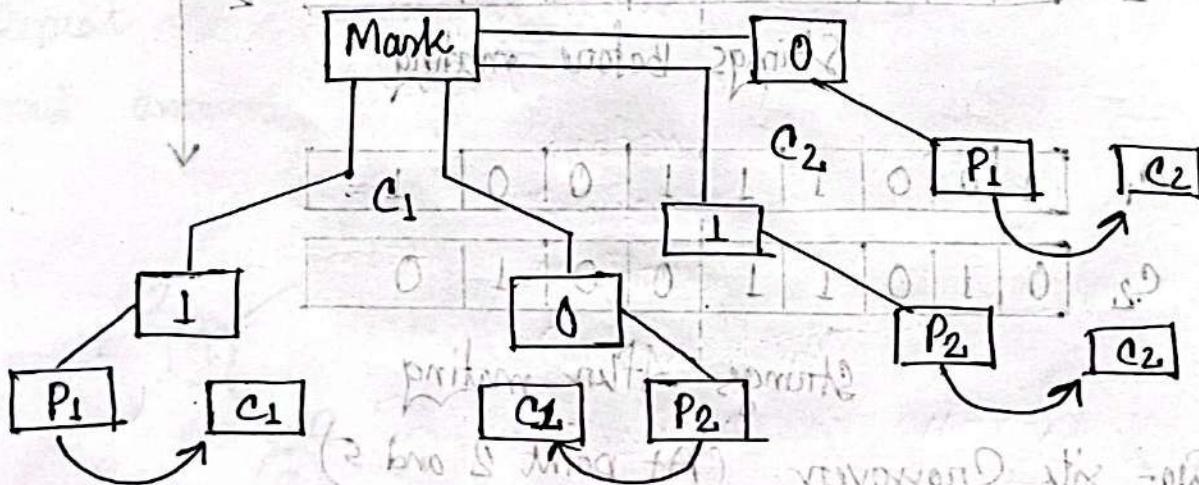
Strings Before mating

c₁ 1 1 0 0 1 1 0 0 0

9 0 0 0 1 1 0 0 1 1

Strings After mating

Mark 10010000110011001



0 1 0 0 1 0 1 0 1 0

109101010 89

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

1010110110 5

Maximizing Function

Maximize the function $f(x) = x^2$, Where the range of x is $[0, 3]$

and maximum number of generation is 3

Step 1: Initialization

00000, 00001, 00010, 00011, 00100, 00101, 00110, 00111
map: number of generation = 30111

Step 2: Selection (1st generation)

Chromosome	Encoding	Value of X	Fitness n^2
C1	00101	5	25
C2	10000	56	256
C3	01001	9	81
C4	11100	28	784
			$\Sigma f(n) = 1146$

Step 3: Crossover (2nd generation)

Chromosome	Encoding	Value of X	χ^2	After Crossover
C ₁	00101	4	16	00100
C ₂	10000	17	289	10001
C ₃	01001	12	144	01100
C ₄	11100	25	625	11001
$\Sigma f(x) = 1074$				

Mutation (and generation)

Chromosome	Encoding	Value of χ^2	$\sum f(x) = n^k$	After Mutation
C ₁	00100	16	4	00100
C ₂	10001	441	21	10101
C ₃	01100	144	12	01100
C ₄	11001	729	27	11011

$(\text{Mutation}) \sum f(x) = 1330, \text{ taken } 2 : 2 \text{ parts}$

Pseudo-Code of Genetic Algorithm

START

Generate the initial population

Compute fitness

REPEAT

Selection

(crossover and mutation : 8 parts)

Mutation

Compute fitness

Initial Data	χ^2	$\sum f(x)$	Final Data
00100	16	4	10100
10001	441	21	00001
01100	144	12	10010
11001	729	27	00111

UNTIL population has not converged

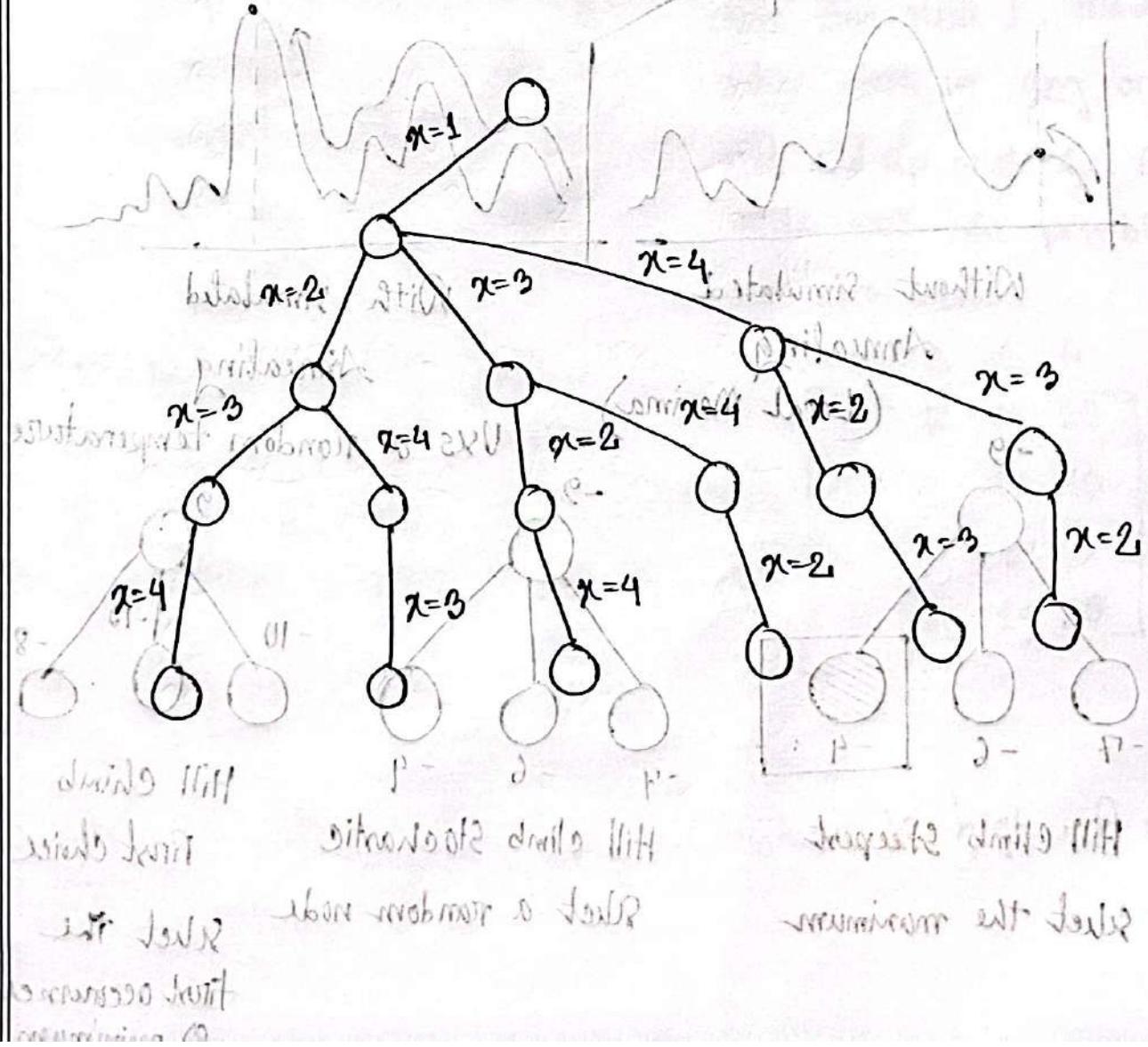
STOP

$\text{PPO} = 0.73$

Queen Problem using Backtracking

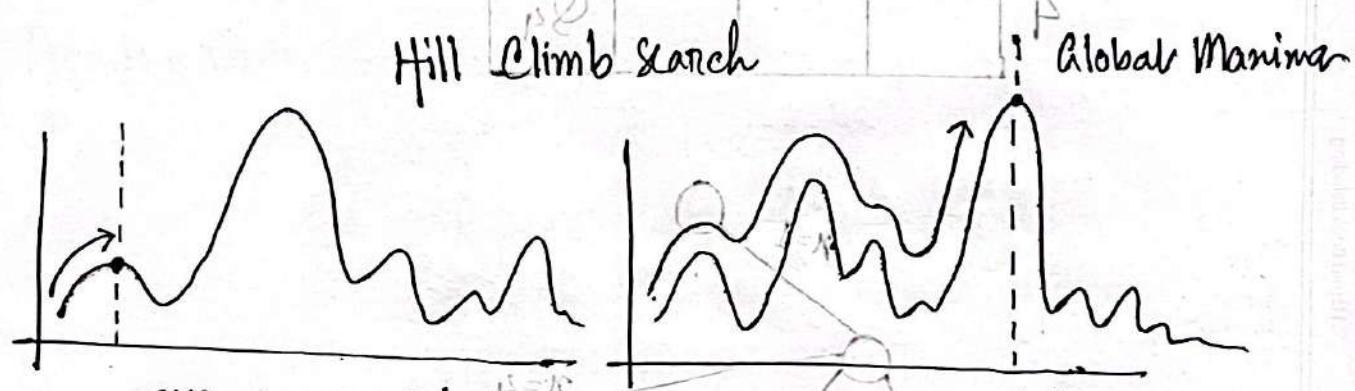
	1	2	3	4
1	Q ₁			
2		Q ₂		
3			Q ₃	
4				Q ₄

Queens cannot be in same row and same columns.



Local Search

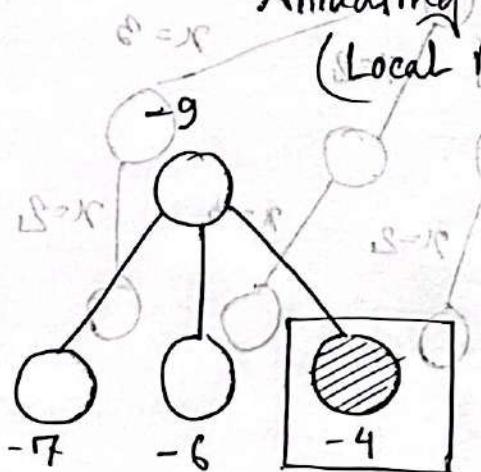
- ① Not concerned with path
- ② Solution itself matters
- ③ Not concerned with perfect solution
- ④ less memory required as its not using Backtracking
- ⑤ less time required
- ⑥ Applicable for large size of graph



Without simulated

Annealing

(Local Maxima)

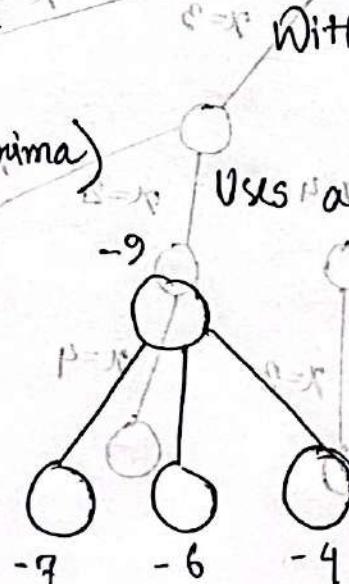


Hill climb steepest
select the maximum

With simulated

Annealing

Uses a random temperature



Hill climb stochastic
select a random node

Hill climb

first choice

select the
first occurrence
or minimum -

Simulated Annealing

It allows downward step also.

Advantages:

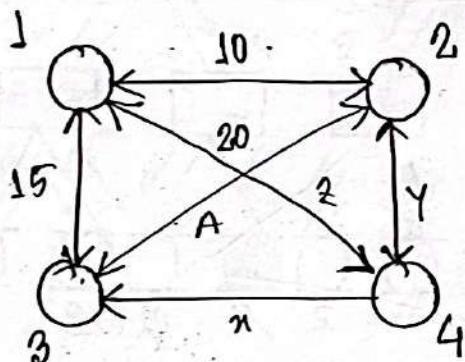
Always gives optimal solutions.

Disadvantages:

Slow process

Can't tell if an optimal solution is found.

TSP (Travelling Salesman Problem)



Start from node 1, travel the entire path in any order and end up with the same node with the possible minimum cost.

$$\Delta E = \text{New Cost} - \text{Old Cost}$$

if $\Delta E < 0$, accept

if $\Delta E \geq 0$, find the probability

$$P = e^{-\Delta E/T}$$

Higher temperature allows worse

Solutions to be accepted. helping

escape local optima. When the temperature is high, P is close to 1.0.

	1	2	3	4
1	0	10	15	20
2	5	0	25	10
3	15	30	0	5
4	15	10	20	0

Question Pattern

Polynomial Problem

Travelling Salesman problem

Local Search Algorithm

Hill Climbing Algorithm

N- Queens Problem

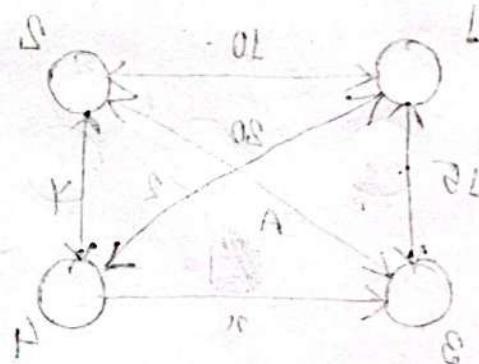
Hill Climbing

Local Search

8-Puzzle / 15-Puzzle Problem

(mildlost viemawal pikkusit) 92T

visit 1 above most frust
no pos in step entro
is left step que bin bin
where left Alice above
this minimum



P Q R S T

08	21	01	0	!
01	28	0	2	8
2	0	08	21	6
0	08	01	21	1

Step 610 - Step 601 = 3Δ

Step 600 - 053Δ7i

Utilizing left bin + 023Δ7i

7Δ6i - 3 = 9

total 2 walls untagged result

Right - between 2nd 3rd row 102

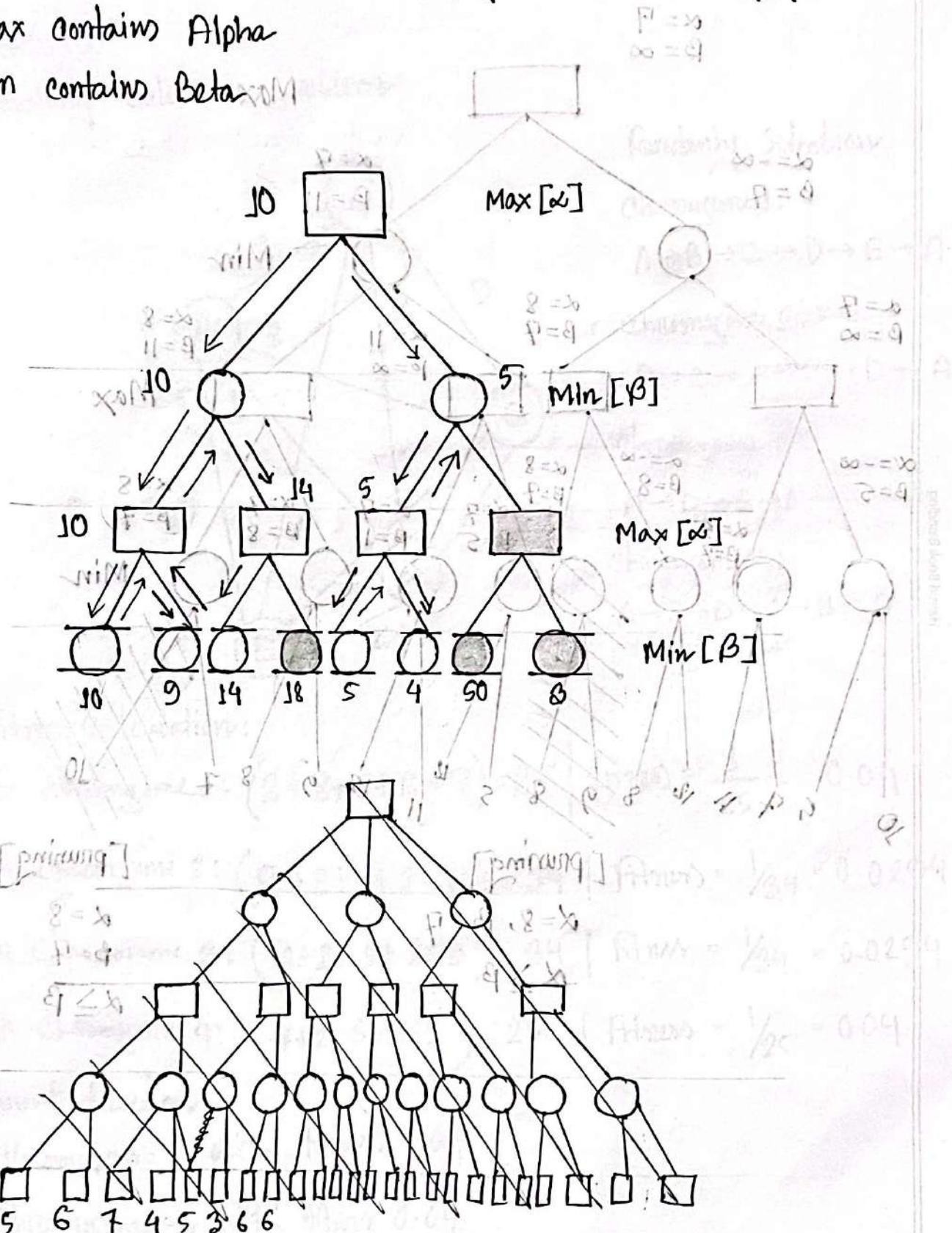
of walls 9, right 21 untagged left wall . smityo last pos 23

• untagged walls

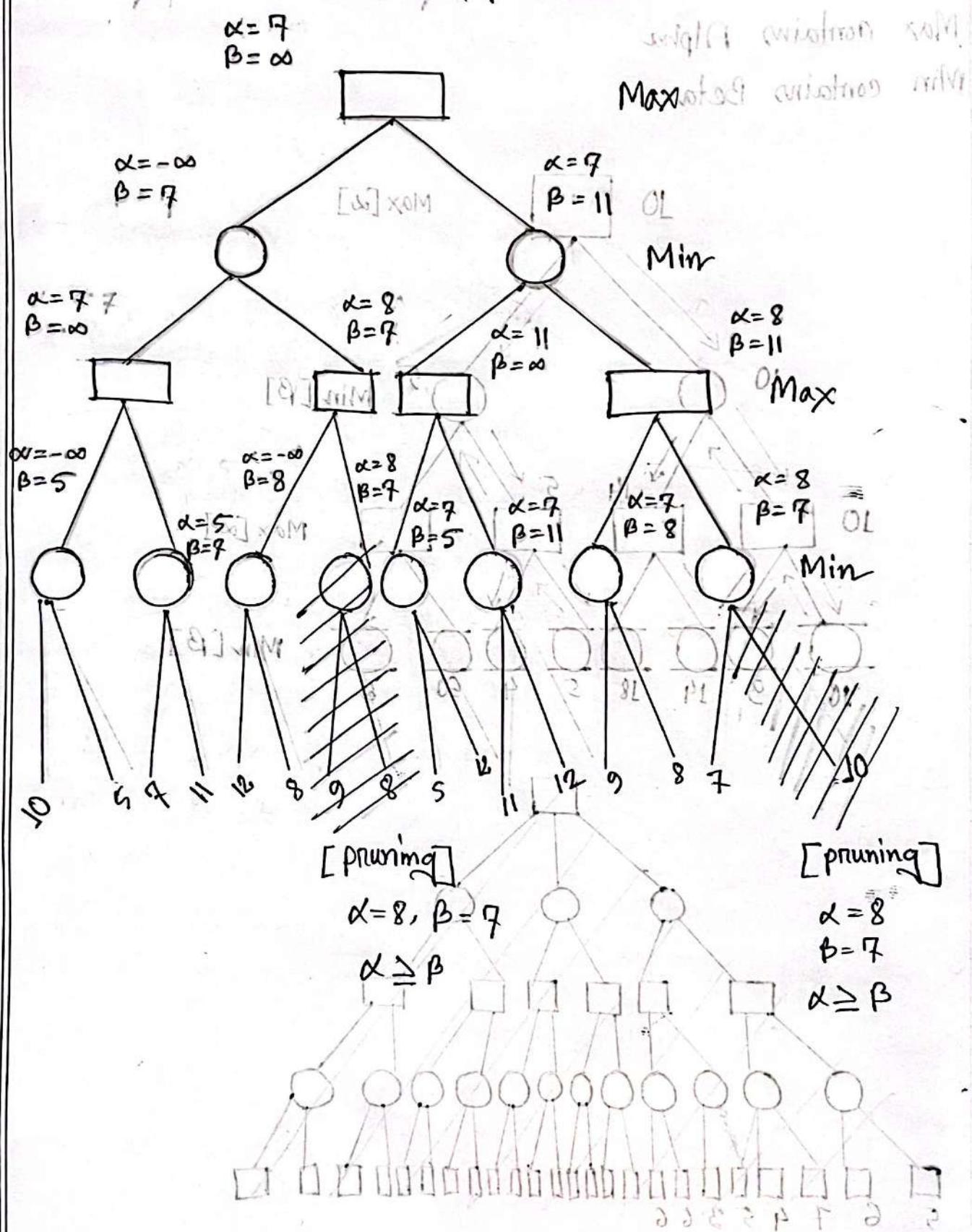
Minimax using Alpha-Beta Pruning

Max contains Alpha

Min contains Beta



Minimax Alpha Beta pruning Simulation



Genetic Algorithm Problem Solving AI

Problem 1:

Travelling Salesman Problem



: Randomly selection:

Chromosome 1:

$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow A$

Chromosome 2:

$A \rightarrow C \rightarrow B \rightarrow E \rightarrow D \rightarrow A$

Chromosome 3:

$A \rightarrow D \rightarrow E \rightarrow B \rightarrow C \rightarrow A$

Chromosome 4:

$A \rightarrow E \rightarrow D \rightarrow C \rightarrow B \rightarrow A$

fitness Calculation:

$$\text{for Chromosome 1: } (2+8+6+2+7) = 25 \quad | \quad \text{Fitness} = \frac{1}{25} = 0.04$$

$$\text{for Chromosome 2: } (9+8+5+2+10) = 34 \quad | \quad \text{Fitness} = \frac{1}{34} = 0.0294$$

$$\text{for Chromosome 3: } (10+2+5+8+9) = 34 \quad | \quad \text{Fitness} = \frac{1}{34} = 0.0294$$

$$\text{for Chromosome 4: } (7+2+6+8+2) = 25 \quad | \quad \text{Fitness} = \frac{1}{25} = 0.04$$

Parent Selection:

Chromosome 1 with fitness 0.04

Chromosome 4 with fitness 0.04

Crossover: IA privata maldon multistopla situm

chromosome 1: A → B → C → D → E → A

chromosome 2: A → E → D → C → B → A

child 1: A → B → D → C → B → A

child 2: A → E → C → D → E → A

Mutation:

: Swap B with E. for ch1

A → E → D → C → B → A | fitness 0.040

Swap C with B for ch2

A → E → B → D → E → A | fitness 0.0416

The answer A → E → B → D → E → A

Distance: 24

$P_{0.0} = \frac{1}{24}$ | $24 = (F + S + D + E + B)$: Lengtgh of route

fitness: 0.0416

$P_{0.0} = \frac{1}{24}$ | $24 = (F + S + D + E + B)$: Lengtgh of route

$P_{0.0} = \frac{1}{24}$ | $24 = (E + S + D + C + B)$: Lengtgh of route

$P_{0.0} = \frac{1}{24}$ | $24 = (S + E + D + C + B)$: Lengtgh of route

$P_{0.0} = \frac{1}{24}$ | $24 = (S + E + D + C + B)$: Lengtgh of route

$P_{0.0} = \frac{1}{24}$ | $24 = (S + E + D + C + B)$: Lengtgh of route

Problem 2:

Rubik's Cube Problem

given that, the final state is

	8	1	6	→ sum = 15
sum 15 ←	3	5	7	-21 +21 -21 +21 -21 21
	4	0	2	21 -21 +21 -21 +21 -21 21
	↓	↓	↓	8 = AF-21 + 21-21 : Chromosome

Chromosome 1:

2	7	6
9	5	1
4	3	8

4	9	2
3	5	7
8	1	6

6	1	8
7	5	3
2	9	4

9	3	5
4	1	8
7	2	6

Fitness calculation:

$$\text{fitness} = \frac{1}{1 + |(15 - \text{row sum})| + |(15 - \text{column sum})| + |(15 - \text{diagonal sum})|}$$

$$1 + |(15 - \text{row sum})| + |(15 - \text{column sum})| + |(15 - \text{diagonal sum})|$$

Chromosome 1:

$$\text{Rows: } |15-15| + |15-15| + |15-15| = 0$$

$$\text{Columns: } |15-15| + |15-15| + |15-15| = 0$$

$$\text{Diagonal: } |15-15| + |15-15| = 0$$

$$\text{fitness} = \frac{1}{(0+1)} = 1$$

Subject:

Date:

Chromosome 2:

fitness: 1

Chromosome 3:

fitness: 1.0

Chromosome 4:

Rows: $|15-17| + |15-13| + |15-15| = 4$ Columns: $|15-16| + |15-13| + |15-16| = 6$ Diagonals: $|15-16| + |15-17| = 3$

fitness: 0.0833

Parent 1: Chromosome 2

Parent 2: Chromosome 1

Child 1:

4	9	2
9	5	1
8	1	6

4	9	2	ch2	ch1
8	0	3	5	7
9	5	1	6	4
8	1	6	3	8

Child 2:

2	7	6
---	---	---

4	3	8
---	---	---

Swap 9 and 3

swap 7 and 9

Mutation

4	3	2
---	---	---

9	5	1
---	---	---

8	1	6
---	---	---

2	9	6
---	---	---

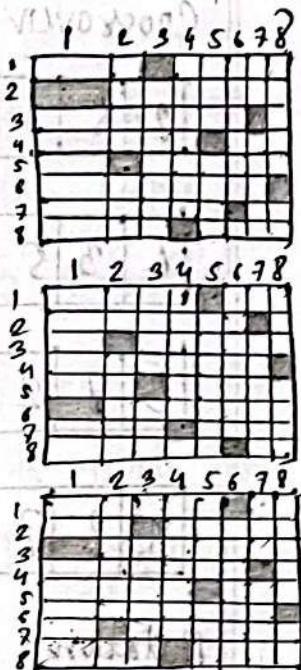
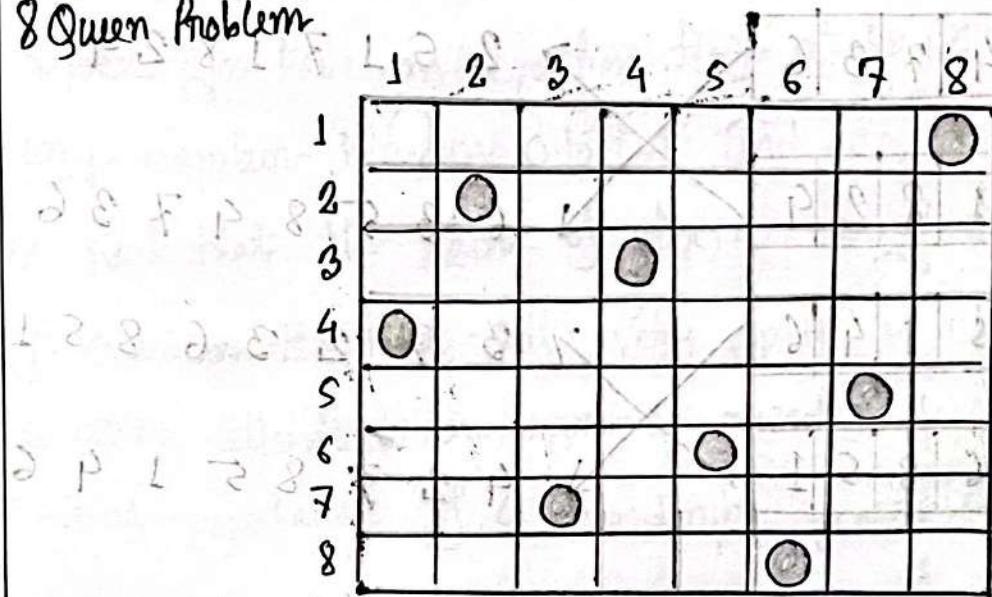
3	5	7
---	---	---

3	3	8
---	---	---

 $0 = |21-21| + |21-21|$
 $\mu = \frac{1}{(1+0)} = 0.5$

Problem 3:

8 Queen Problem



Representation:

$[4, 2, 7, 3, 6, 8, 5, 1]^8$ Here, index = column
Value = row

Chromosome 1: $[4, 2, 7, 3, 6, 8, 5, 1]$

Chromosome 2: $[2, 5, 1, 8, 4, 7, 3, 6]$

Chromosome 3: $[6, 3, 5, 7, 1, 8, 2, 4]$

Chromosome 4: $[3, 7, 2, 8, 5, 1, 4, 6]$

fitness:

$$\begin{array}{llll}
 q_1 = 6 & \frac{\text{total}}{56} & q_1 = 8 & \frac{\text{total}}{64} \\
 q_2 = 8 & \frac{56}{56} & q_2 = 8 & \frac{64}{64} \\
 q_3 = 6 & \Sigma T = 240 & q_3 = 8 & \Sigma T = 240 \\
 q_4 = 8 & & q_4 = 8 & \\
 q_5 = 8 & \text{fitness} & q_5 = 8 & \text{fitness} \\
 q_6 = 6 & 23.33\% & q_6 = 8 & 35\% \\
 q_7 = 8 & & q_7 = 8 & \\
 q_8 = 6 & & q_8 = 8 & \\
 & & & q_8 = 8
 \end{array}$$

Crossover:

2	5	1	8	4	7	3	6
---	---	---	---	---	---	---	---

6	3	5	7	1	8	2	4
---	---	---	---	---	---	---	---

3	7	2	8	5	1	4	6
---	---	---	---	---	---	---	---

4	2	7	3	6	8	5	1
---	---	---	---	---	---	---	---

2	5	1	7	3	8	2	4
---	---	---	---	---	---	---	---

6	3	5	8	4	7	3	6
---	---	---	---	---	---	---	---

3	7	2	3	6	8	5	1
---	---	---	---	---	---	---	---

4	2	7	8	5	1	4	6
---	---	---	---	---	---	---	---

Mutation:

2	5	1	8
---	---	---	---

6	3	5	8	4	4	3	6
---	---	---	---	---	---	---	---

3	7	2	1	6	8	5	1
---	---	---	---	---	---	---	---

4	2	1	8	8	1	4	6
---	---	---	---	---	---	---	---

1	8	2	4	3	5	6
---	---	---	---	---	---	---

2	8	4	4	3	6
---	---	---	---	---	---

2	8	4	4	3	6
---	---	---	---	---	---

2	8	4	4	3	6
---	---	---	---	---	---

2	8	4	4	3	6
---	---	---	---	---	---

2	8	4	4	3	6
---	---	---	---	---	---

Jobot

8 = N

Jobot

8 = N

Jobot

8 = N

Jobot

8 = N

D

8 = N

P

8 = N

D

8 = N

D

8 = N

OP

8 = N

OP

8 = N

OP

8 = N

OP

8 = N

T3

8 = N

OP

8 = N

OP

8 = N

OP

8 = N

WTF

8 = N

WTF

8 = N

WTF

8 = N

WTF

8 = N

PES

8 = N

PES

8 = N

PES

8 = N

PES

8 = N

P

8 = N

P

8 = N

P

8 = N

P

8 = N

E

8 = N

E

8 = N

E

8 = N

E

8 = N

S

8 = N

S

8 = N

S

8 = N

S

8 = N

Mid Question (Genetic Algorithm)

1. Suppose you have an equation $f(x) = x^2 - 5x + 6$. Assume x can be any number between 0 to 15. Find an appropriate value of x such that the value of $f(x) = 0$ using Genetic Algorithm
- (a) Consider the fact that every population chromosome will have 4 genes. Illustrate an appropriate encoding system to create an initial population of 4 randomly generated chromosome.

$$f(x) = x^2 - 5x + 6$$

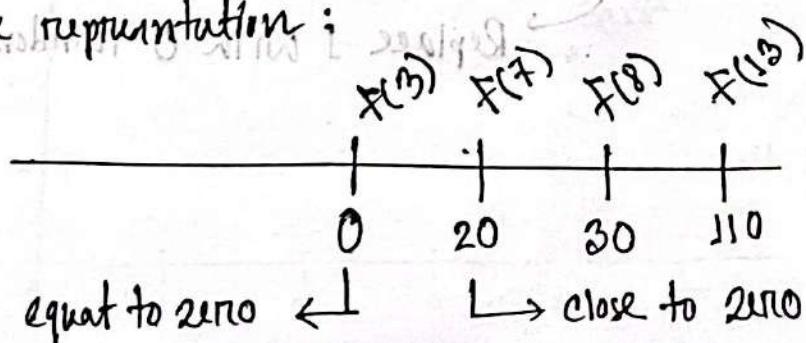
BCD representation: (Encoding system)

1	1	0	8	4	2	1	0
0	1	1	1	1	1	1	1
0	0	1	1	1	1	1	1
1	0	0	0	0	0	0	0
1	1	0	0	1	1	1	1

chromosome 1
chromosome 2
chromosome 3
chromosome 4

- (b) Using an appropriate fitness function deduce the 2 fittest chromosome and perform a single point crossover from the middle to create two offspring.

Number Line representation:



parent 1: fitness ∞

0	0	1	1
---	---	---	---

parent 2: fitness 0.05

0	1	1	1
---	---	---	---

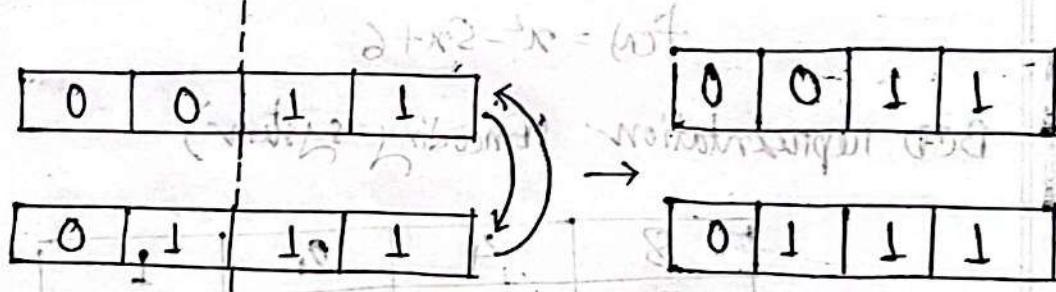
fitness calculation

$$\text{fitness} = \frac{1}{f(x)}$$

for $f(3)$, fitness = ∞ for $f(7)$, fitness = 0.05for $f(8)$, fitness = 0.0333for $f(13)$, fitness = 0.09×10^{-3}

single point crossover:

Assume, point = 2.



c) Explain how you can mutate the offspring.

Mutation

Replace 0 with 1 randomly

1	0	1	1
---	---	---	---

$$\text{fitness: } \frac{1}{f(11)} = 0.01388$$

0	0	1	1
---	---	---	---

$$\text{fitness: } 0.05 \infty$$

Replace 1 with 0 randomly

1	1	1	1
---	---	---	---

0.1% of 20/20 \leftarrow 1 \rightarrow 0.05% of 20/20

Q) Explain your opinion on whether Genetic Algorithm can be treated as a class of Local Search Algorithms or not.

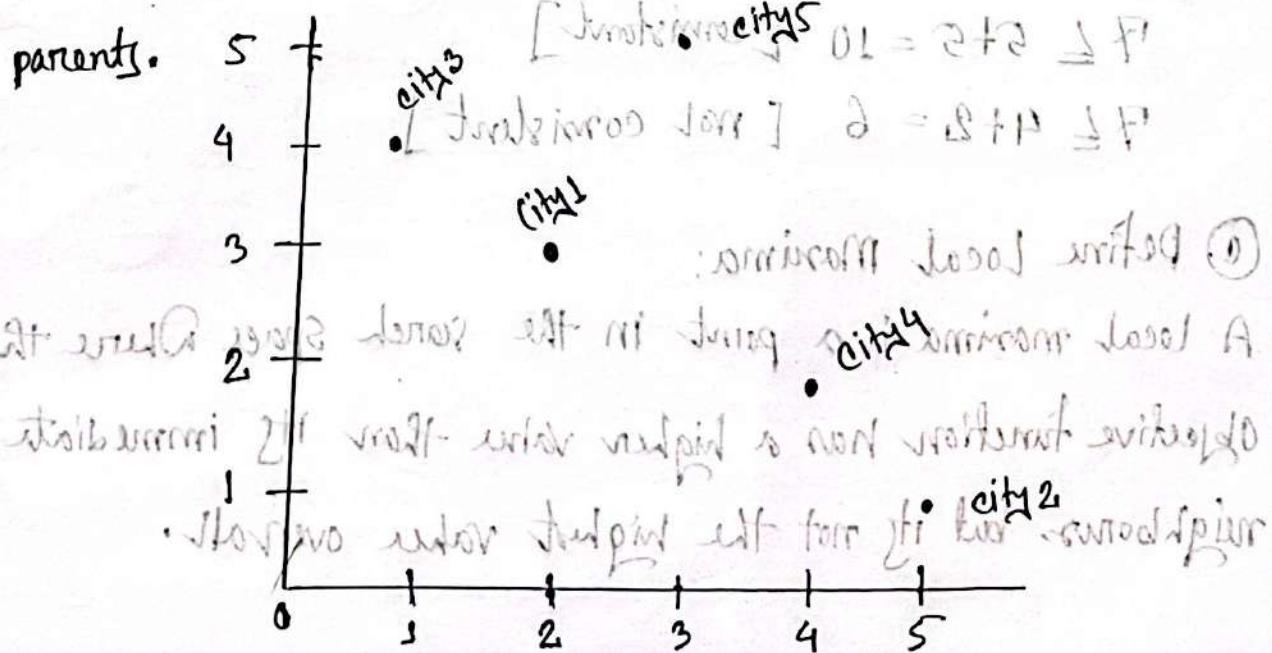
GA is better to be considered as global search algorithm for its diversity and mutation nature.

2. Ruhani is a sales person. His point is to find the optimal path that visits all cities at once and return to the starting point.

City 1	City 2	City 3	City 4	City 5
(2,3)	(5,1)	(1,4)	(4,2)	(3,5)

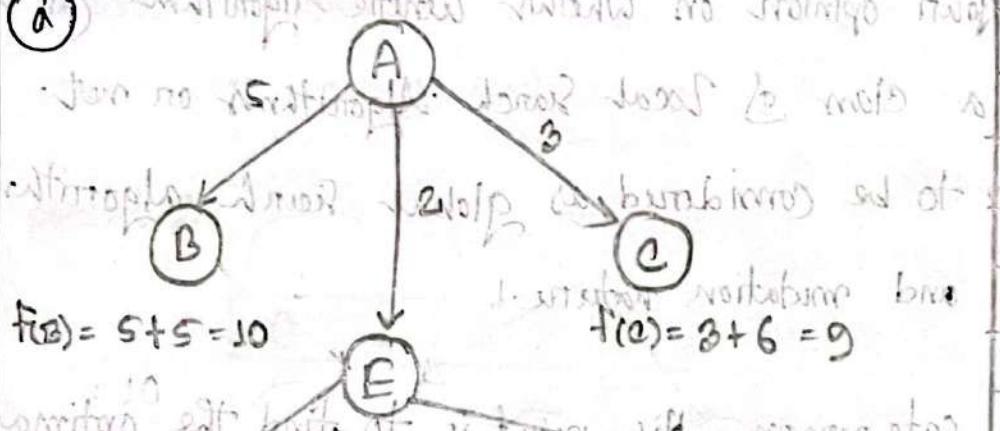
Use genetic algorithm to find solutions.

Q) Encode the problem and create four parent chromosomes. Then determine an appropriate fitness function and choose parents.

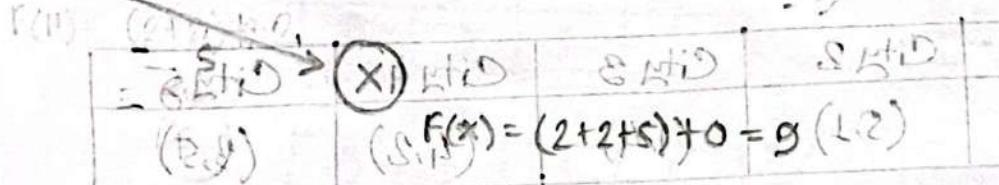
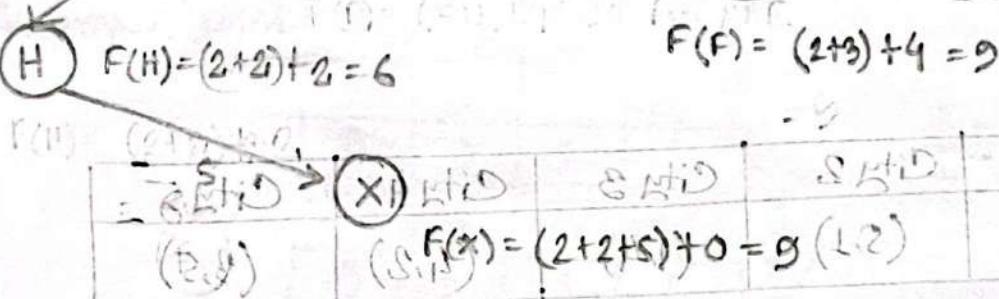


3.

(a)



Node	h-value
A	7
B	5
C	6
D	3
E	4
F	4
G	3
H	2
I	0

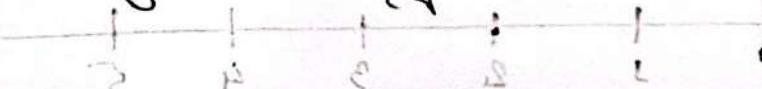


(b) Check the Consistency.

Heuristic value of Node 1 \leq Heuristic value of node 2 + path costNode A: $7 \leq 5+5 = 10$ [consistent] $7 \leq 4+2 = 6$ [not consistent].

(c) Define Local Maxima:

A local maxima is a point in the search space where the objective function has a higher value than its immediate neighbors, but it's not the highest value overall.



Example: Hill Climbing. it will stuck at local maxima but there will be a better solution elsewhere.

③ Define local minima:

A local minima is a point where the objective function has a lower value than its immediate neighbour but not the lowest at all.

Hill climbing: often stuck in local maxima

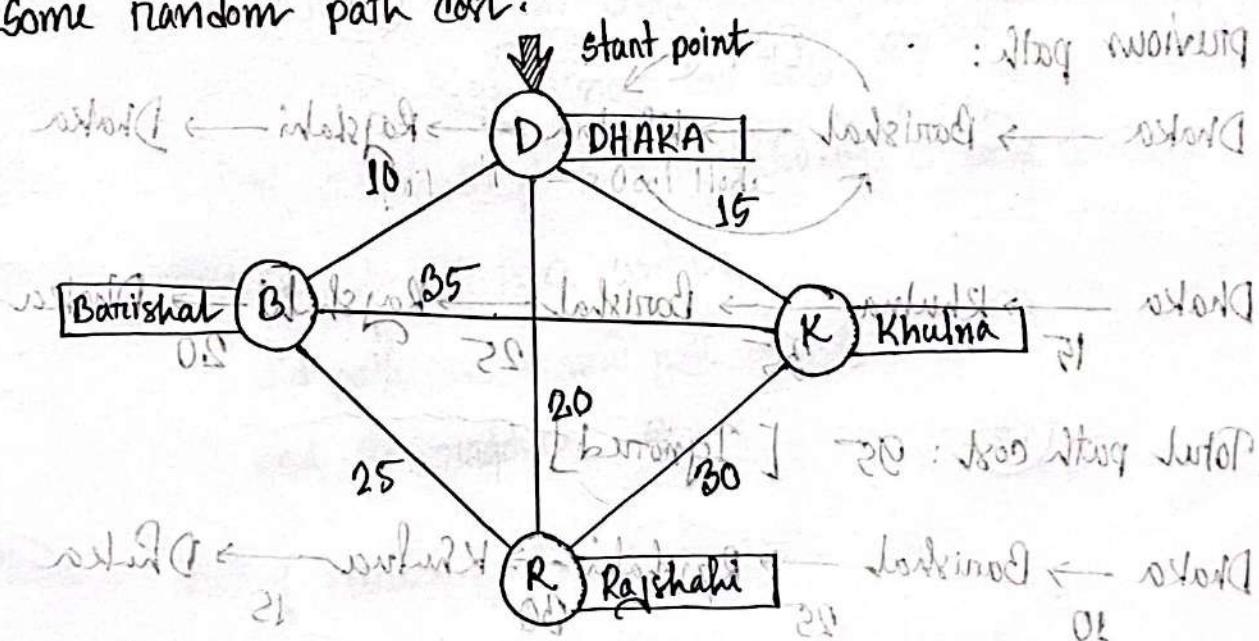
Simulated Annealing: Escape local maxima/minima

GA: Avoid local traps and find global optima

Question 2:

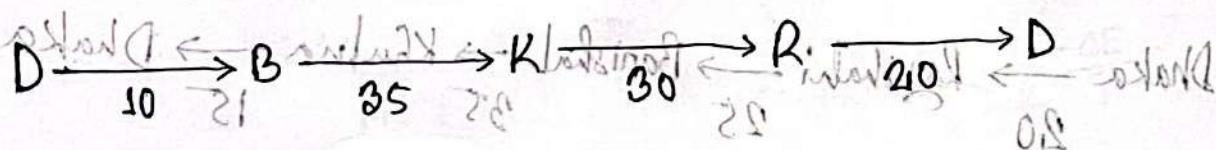
Travelling Salesman Problem is a path travelling problem where the algorithm finds that path which is the shortest among all other paths possible from start point to start point after visiting every other nodes exactly one time.

To solve the problem let assume that we have 4 cities and they are interconnected as mesh topology structure with some random path cost.

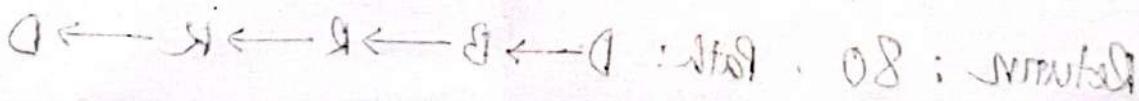


Approach:

① Select any initial Route:



Total distance = 95 unit



⑥ Generate Random path:

We will generate random path by swapping any two cities. Then we will calculate the path cost.

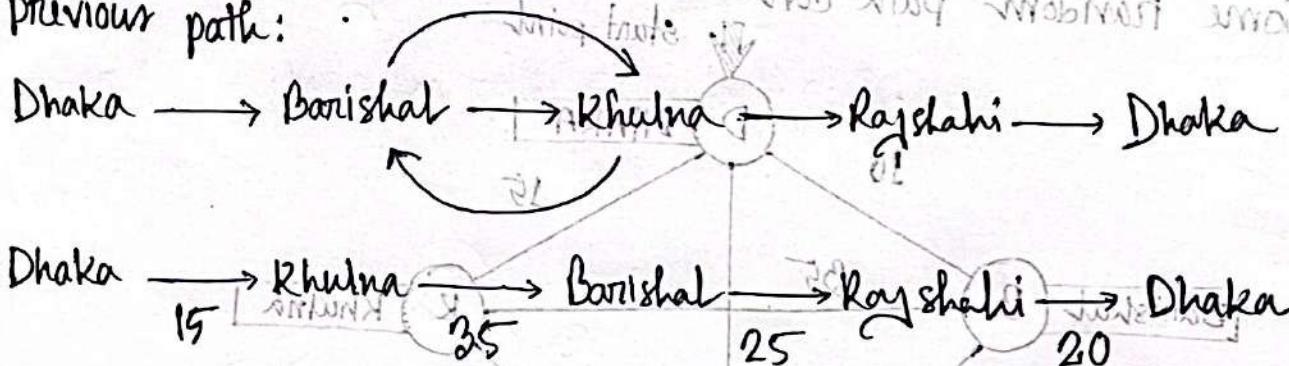
If the total cost \leq previous cost:

Ignore!

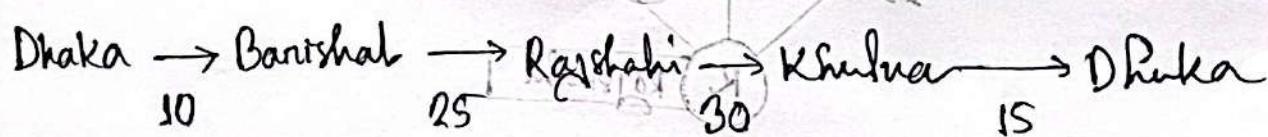
If the total cost $<$ previous cost:

Accepted!

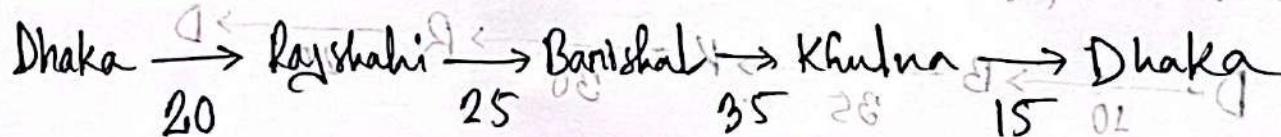
previous path:



Total path cost: 95 [Ignored]



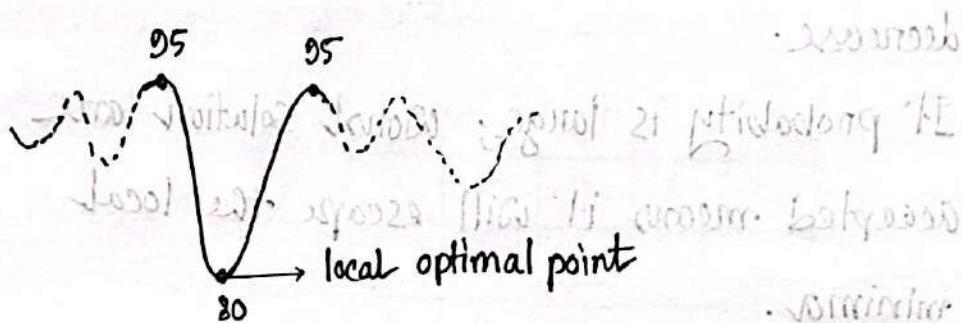
Total path cost: 80 [Accepted]



Total path cost: 95 [Ignored and Stop searching]

Return: 80 , Path: D → B → R → K → D

Visual Representation:



Question 3:

- Step A: Initialize a random path and find the path cost.
- Step B: Initialize a random temperature. The selected temperature neither too high nor too low.
- Step C: Define a cooling factor between 0 and 1.
- Step D: Repeat step A, B, C until:
- old path cost - new path cost ≤ 0
and temperature value is very small.
- Step E: Every iteration / function call
- compute $\Delta F = F(S_{\text{new}}) - F(S_{\text{old}})$
if $\Delta F \leq 0$; ACCEPTED
else:
ACCEPTED with probability $= e^{-\frac{\Delta F}{T}}$
 $T = \alpha$ (cooling factor) $\times T$

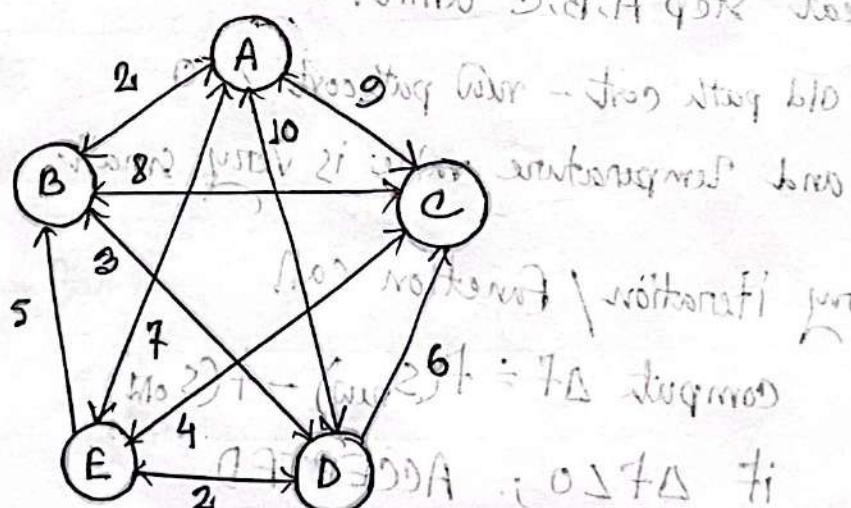
Step F: Every iteration the temperature (T) will decrease.

If probability is large; worst solution are accepted means it will escape the local minima.

Step G: When the T is very low to make the next search, the program will stop finding.

Step H: Return the path with path cost.

Question 4:



$T = 2$ - After 4379700A

$$P = \min(1, e^{-\frac{f_c - f_b}{kT}}) \quad \text{So } P = 1$$

Iteration 1:

Initialization of Chromosome:

Chromosome 1:

A	B	C	D	E	A
---	---	---	---	---	---

Chromosome 2:

A	C	B	E	D	A
---	---	---	---	---	---

Chromosome 3:

A	D	E	B	C	A
---	---	---	---	---	---

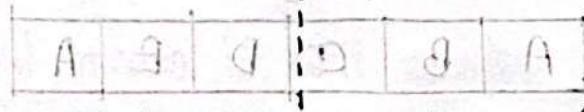
Chromosome 4:

A	E	D	C	B	A
---	---	---	---	---	---

Fitness Calculations:

For Chromosome 1: $(2+8+6+2+7) = 25$ Fitness: $\frac{1}{25} = 0.04$ [Highest fitness]For Chromosome 2: $(9+8+5+2+10) = 34$ Fitness: $\frac{1}{34} = 0.0294$ For Chromosome 3: $(10+2+5+8+9) = 34$ Fitness: $\frac{1}{34} = 0.0294$ For Chromosome 4: $(9+2+6+8+2) = 25$ Fitness: $\frac{1}{25} = 0.04$ [Highest fitness]

Crossover:

chromosome 1: $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow A$: L crossoverchromosome 2: $A \rightarrow E \rightarrow D \rightarrow C \rightarrow B \rightarrow A$: I crossoverchild 1: $A \rightarrow B \rightarrow D \rightarrow C \rightarrow B \rightarrow A$: S crossoverchild 2: $A \rightarrow E \rightarrow C \rightarrow D \rightarrow E \rightarrow A$

Mutation:

swap B with E for child 1 $A | A$

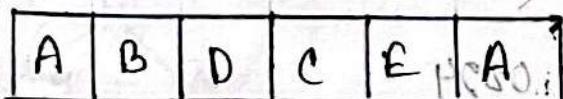
swap E with B for child 2

child 1: $A \rightarrow B \rightarrow D \rightarrow C \rightarrow EA \rightarrow A$

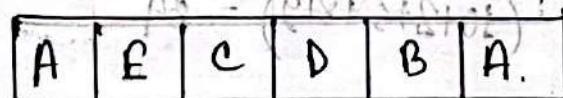
child 2:

 $A \rightarrow E \rightarrow C \rightarrow D \rightarrow B \rightarrow A$: mutation $P_m = (P_1 + P_2 + P_3 + P_4) / 4$: L crossover not

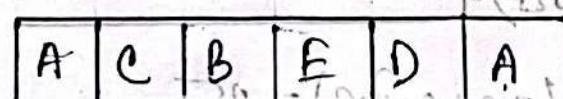
Iteration 2:

Initialize chromogome: $P_m = \frac{1}{20}$: mutationchromosome 1: $P_m = (0.1 + 0.2 + 0.3 + 0.4) / 4$: S crossover not

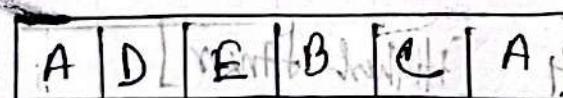
chromosome 2:



chromosome 3:



chromosome 4:



child 1 and
child 2 in
previous iteration
 $P_m = \frac{1}{20}$: mutation

for child 3 and
 $P_m = \frac{1}{20}$: mutation

Fitness Calculations:

for chromosome 1: $(2+3+6+4+7) = 22$

$$\text{fitness} = \frac{1}{22} = 0.04545 \quad [\text{Highest fitness}]$$

for chromosome 2: $(7+4+6+3+2) = 22$

$$\text{fitness} = \frac{1}{22} = 0.04545 \quad [\text{Highest fitness}]$$

for chromosome 3: $(0+8+5+2+10) = 34$

$$\text{fitness} = 0.029411$$

for chromosome 4: $(10+2+5+8+9) = 34$

$$\text{fitness} = \frac{1}{34} = 0.029411$$

Crossover:

A	0	0	0	0	A
---	---	---	---	---	---

chromosome 1: $A \rightarrow B \rightarrow D \rightarrow C \rightarrow E \rightarrow A$

chromosome 2: $A \rightarrow E \rightarrow C \rightarrow D \rightarrow B \rightarrow A$

Child 1: $A \rightarrow B \rightarrow D \rightarrow D \rightarrow B \rightarrow A$

Child 2: $A \rightarrow E \rightarrow C \rightarrow C \rightarrow E \rightarrow A$

Mutation:

Swap D with C for child 1. $0 = p_1^M = \text{mut1}$

Swap C with B for child 2. $11P(0.0) = p_2^M = \text{mut2}$

child 1: $A \rightarrow B \rightarrow C \rightarrow D \rightarrow B \rightarrow A$ correct

Child 2: $A \rightarrow B \rightarrow B \rightarrow C \rightarrow E \rightarrow A$ not correct

Iteration 3:

Initialize chromosomes

Chromosome 1:

A	B	C	D	B	A
---	---	---	---	---	---

chromosome 2:

A	E	B	C	E	A
---	---	---	---	---	---

chromosome 3:

A	C	B	E	D	A
---	---	---	---	---	---

chromosome 4:

A	D	E	B	C	A
---	---	---	---	---	---

fitness Calculations:

for chromosome 1: $(2+8+6+3+2) = 21$ correct

fitness = $\frac{1}{21} = 0.04761$ [Highest fitness]

for chromosome 2: $(7+5+8+4+7) = 31$

fitness = $\frac{1}{31} = 0.03225$ [Highest fitness]

for chromosome 3: $(9+8+5+2+10) = 34$

fitness = $\frac{1}{34} = 0.029411$

for chromosome 4: $(5+2+5+8+9) = 31$

fitness = $\frac{1}{31} = 0.03225$

Crossover:

Chromosome 1: $A \rightarrow B \rightarrow C \xrightarrow{H} D \xrightarrow{O} B \rightarrow A$

Chromosome 2: $A \rightarrow E \rightarrow B \rightarrow C \xrightarrow{H} E \rightarrow A$

Child 1: $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow A$

Child 2: $A \rightarrow E \rightarrow B \rightarrow C \rightarrow B \rightarrow A$

Mutations:

swap B with D in child 1

swap B with D in child 2

Child 1: $A \rightarrow D \rightarrow C \rightarrow D \rightarrow E \rightarrow A$

$$\text{fitness} = \frac{1}{(10+6+6+2+7)} = 0.032258$$

Child 2: $A \rightarrow E \rightarrow B \rightarrow C \rightarrow D \rightarrow A$

$$\text{fitness} = \frac{1}{(7+5+8+6+10)} = 0.032258$$

$A \rightarrow E \rightarrow D \rightarrow C \rightarrow B \rightarrow A$

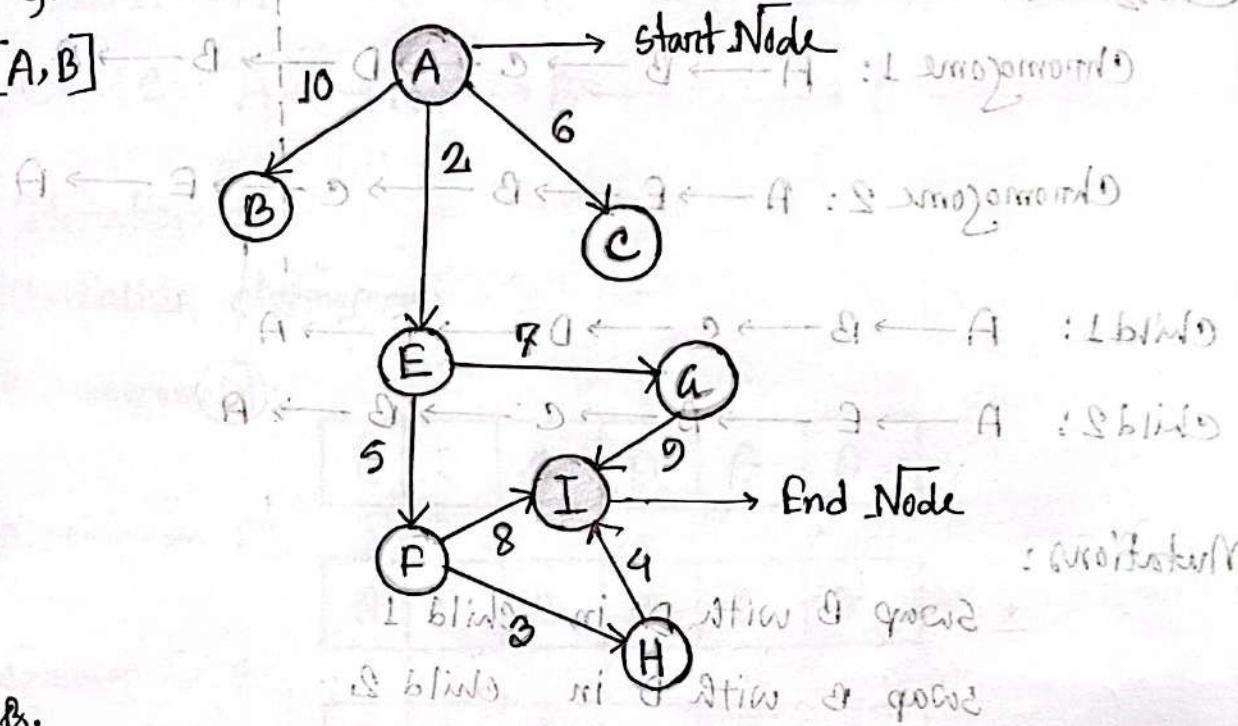
$$\text{fitness} = \frac{1}{(7+2+6+8+2)} = \frac{1}{25} = 0.04$$

Final Path: $A \rightarrow E \rightarrow D \rightarrow C \rightarrow B \rightarrow A$

Cost: 25, fitness: 0.04.

Question 1:

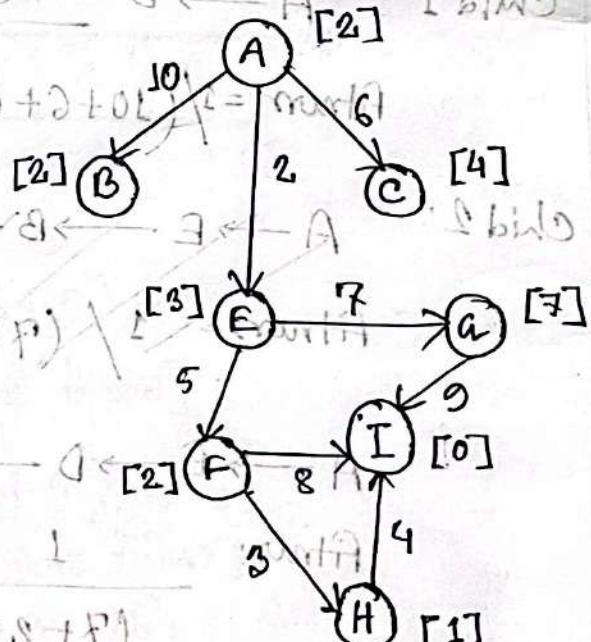
[A, B]



B.

C.

Node	Heuristic Value
A	0
B	2
C	4
E	3
F	2
a	7
H	1



D.

Start Node = A

End Node = I

find the lowest / shortest path from A to I

path 1: $A \xrightarrow{2} E \xrightarrow{5} F \xrightarrow{8} I$, cost = 15path 2: $A \xrightarrow{2} E \xrightarrow{7} G \xrightarrow{9} I$, cost = 18path 3: $A \xrightarrow{2} E \xrightarrow{5} F \xrightarrow{3} H \xrightarrow{4} I$, cost = 14∴ The shortest path: $A \rightarrow E \rightarrow F \rightarrow H \rightarrow I$

The graph is admissible if Node A, E, F, H and I are admissible simultaneously.

for checking admissibility,

Admissible if, heuristic value of (n) \leq actual path cost from (n)

for Node A:

$$h(A) = 2 \text{ and } h^*(A) = (2 + 5 + 3 + 4) = 14$$

 $\therefore h(A) \leq h^*(A)$

for Node B:

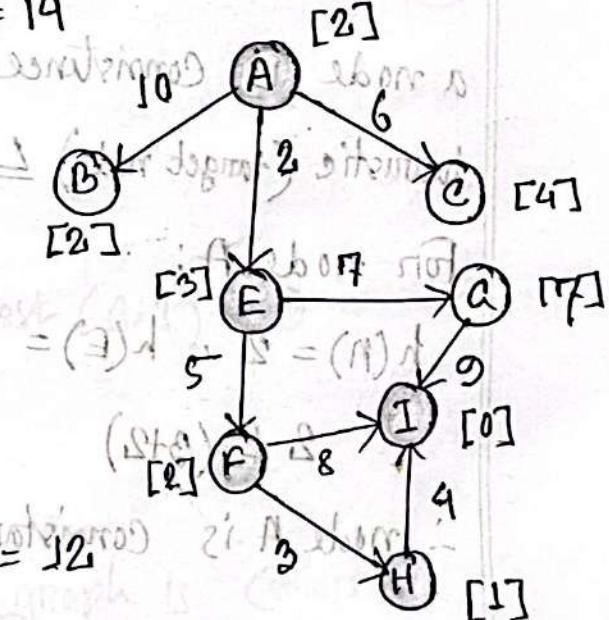
$$h(B) = 2 \text{ [edge node]}$$

for Node C:

$$h(C) = 4 \text{ [edge node]}$$

for Node E:

$$h(E) = 3 \text{ and } h^*(E) = (5 + 3 + 4) = 12$$

 $\therefore h(E) \leq h^*(E)$


for Node F:

$$h(F) = 2 \text{ and } h^*(F) = (3+4) = 7$$

$\therefore h(F) \leq h^*(F)$ ∴ A is a valid heuristic

A = valid heuristic

F = valid node

for Node H:

$$h(H) = 1 \text{ and } h^*(H) = 4$$

$\therefore h(H) \leq h^*(H)$ ∴ A is a valid heuristic

for Node I:

$$h(I) = 0 \text{ and } h^*(I) = 0$$

$\therefore h(I) \leq h^*(I)$ ∴ A is a valid heuristic

$$h(A) = 7 \text{ and } h^*(A) = 9$$

$\therefore h(A) \leq h^*(A)$

\therefore The graph is admissible. [proved]

E. The shortest path from Node A to Node I is

$$A \rightarrow E \rightarrow F \rightarrow H \rightarrow I$$

a node is consistent if

$$\text{heuristic (target node)} \leq \text{heuristic (next node)} + \text{path cost (m, n)}$$

for node A:

$$h(A) = 2, h(E) = 3, \text{ path cost (A, E)} = 2$$

$$2 \leq (3+2)$$

\therefore node A is consistent.

For node E:

$$h(E) = 3, \quad h(F) = 2, \quad \text{path cost}(E, F) = 5$$

$$h(G) = 7, \quad \text{path cost}(E, G) = 7$$

$$3 \leq (2+5) \quad \text{and} \quad 3 \leq (7+7)$$

∴ node E is consistence.

For node F:

$$h(F) = 2, \quad h(I) = 0, \quad \text{path cost}(F, I) = 8$$

$$h(H) = 1, \quad \text{path cost}(F, H) = 3$$

$$2 \leq (0+8) \quad \text{and} \quad 2 \leq (1+3)$$

∴ Node F is consistence.

For node H:

$$h(H) = 1, \quad h(I) = 0, \quad \text{path cost}(I, H) = 4$$

$$1 \leq (0+4)$$

∴ node H is consistence.

For node G:

$$h(G) = 0, \quad h(A) = 0, \quad \text{path cost}(A, G) = 9$$

$$0 \leq (7+9)$$

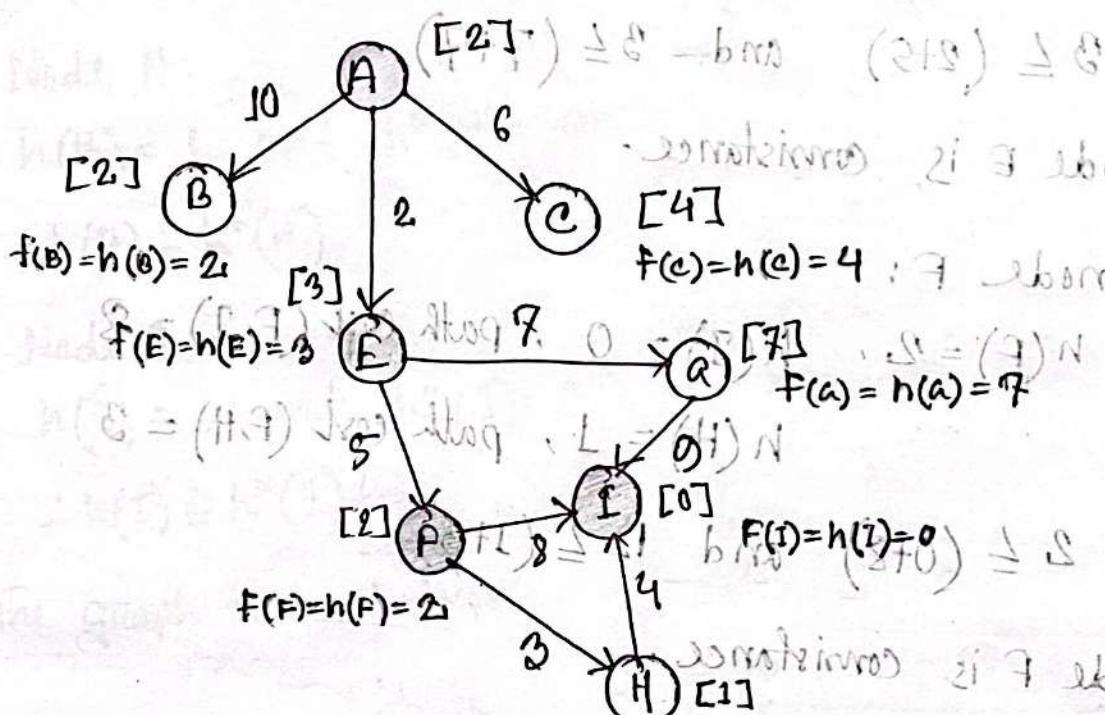
∴ node G is consistence. (For last marking) now answer

As all nodes are consistence. The graph is consistence.

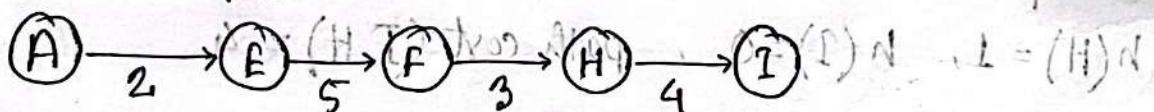
F.

Technique 1: $f = (g, h)$ where $g = f$ & $h = f$

Greedy Best First (Search) where $f = h$

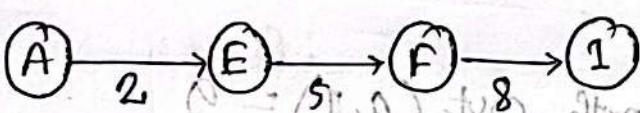


The actual path:



Total path cost = 14

GBFS path:



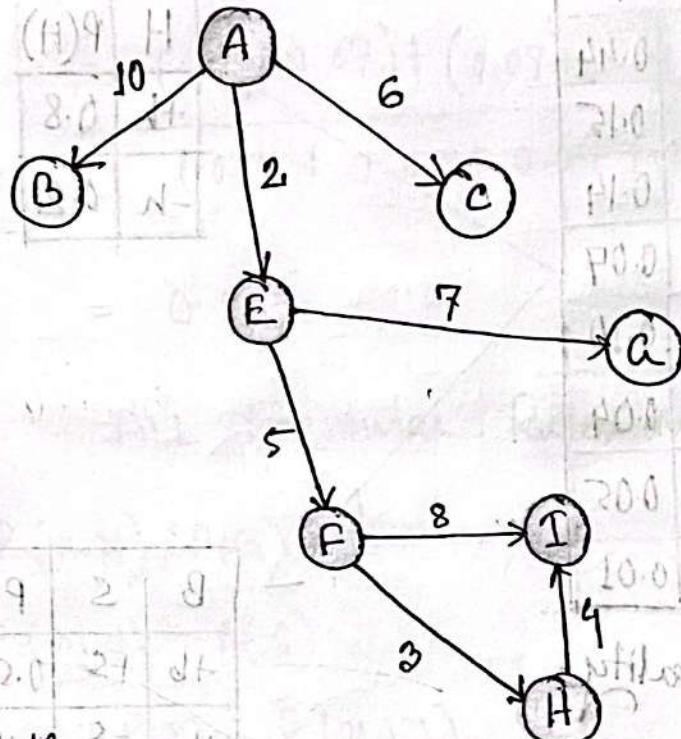
Total path cost = 15

$(E+F) \leq 10$

Conclusion: Completeness but not optimal

Technique 2:

Uniform Cost Search Algorithm



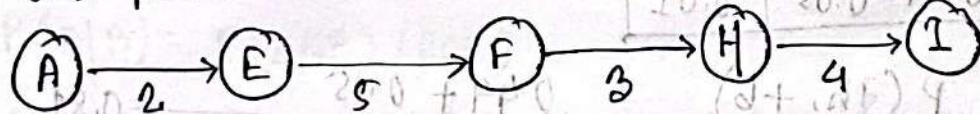
9	2	8	H
P10.0	2+	dt	dt
20.0	2	dt	dt
P10.0	2+	dt	dt
P10.0	2-	dt	dt
10.0	2+	dt	dt
P10.0	2-	dt	dt
20.0	2+	dt	dt
P10.0	2-	dt	dt
20.0	2+	dt	dt
10.0	2-	dt	dt

The actual path:



Total path cost: 14

UCS path:



Total path cost = 14

Conclusion: Complete and optimal simultaneously.

Not always complete and optimal.

Probability

↳ expected

→ simple, basic (no) notation

H	B	S	P
th	+b	+s	0.44
th	+b	-s	0.15
th	-b	+s	0.14
th	-b	-s	0.07
-h	+b	+s	0.10
-h	+b	-s	0.04
-h	-b	+s	0.05
-h	-b	-s	0.01

H	P(H)
th	0.8
-h	0.2

Conditional Probability

	+b	-b		
	+s	-s	+s	-s
th	0.44	0.15	0.14	0.07
-h	0.10	0.04	0.05	0.01

B	S	P
+b	+s	0.54
+b	-s	0.19
-b	+s	0.19
-b	-s	0.02

(a) $P(+h|+b) = \frac{P(+h, +b)}{P(+b)} = \frac{0.44 + 0.15}{(0.44 + 0.15 + 0.1 + 0.04)} = 0.81$

(b) $P(-h, -s|+b) = \frac{P(-h, -s, +b)}{P(+b)} = \frac{0.04}{(0.44 + 0.15 + 0.1 + 0.04)} = 0.05$

(c) $P(-h|+s, -b) = \frac{P(-h, +s, -b)}{P(+s, -b)} = \frac{0.05}{(0.44 + 0.05)} = 0.26$

$$\begin{aligned}
 \textcircled{1} \quad P(+h \vee -s \mid -b) &= \frac{(P(+h, -b) + P(-s, -b) - P(+h, -s, -b))}{P(-b)} \\
 &= \frac{(0.14 + 0.07) + (0.07 + 0.01) - 0.07}{(0.14 + 0.05 + 0.07 + 0.01)} \\
 &= 0.81
 \end{aligned}$$

ALL Formulae, Rules and Principles

$$P(x \mid y) = P(x, y) / P(y)$$

$$P(x, y) = P(x \mid y) * P(y)$$

$$P(x \vee y) = P(x) + P(y) - P(x \wedge y)$$

$$P(x, y) = P(x) * P(y)$$

$$P(x, y \mid z) = P(x \mid z) * P(y \mid z)$$

$$P(A') = 1 - P(A)$$

$$P(A \mid B) = P(A \wedge B) / P(B)$$

$$P(A \wedge B) = P(A) * P(B) \quad \text{if A and B are independent}$$

$$P(A \mid B) = \frac{(P(B \mid A) * P(A))}{P(B)} \quad (\text{Bayes' Theorem})$$

Marginal Probability:

	Male	Female	Total
GOT	0.16	0.24	0.40
TBBT	0.2	0.05	0.25
Other	0.1	0.25	0.35
Total	0.46	0.54	1

(a) What is the marginal probability of people watching TBBT?

$$P(TBBT) = 0.25$$

$$(B|TBBT) = (B|X) = 0.25$$

(b) What is the joint probability of a person being female and liking TBBT?

$$P(F \wedge T) = 0.05$$

$$(B|F) = (B|X) = 0.05$$

(c) What is the probability of a person liking GOT given that person is male?

$$P(GOT | \text{Male}) = \frac{P(GOT \wedge \text{Male})}{P(\text{Male})}$$

$$= \frac{0.16}{0.46} = 0.347826$$

(2)

(method copied)

$$\frac{((A) = 0.46 \times (A|F) = 0.46 \times 0.1)}{(A) = 0.46}$$

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad | \quad P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \begin{array}{l} \text{Joint probability} \\ \rightarrow (A \cap B) \end{array}$$

8P.04 SEP.0

$$\rightarrow \text{Marginal Probability}$$

$$(A|B) = (A \cap B) / (B) \quad (B|A) = (A \cap B) / (A)$$

$$\text{Hence, } P(A|B) \times P(B) = P(B|A) \times P(A)$$

$$\text{Then, } P(A|B) = \frac{(P(B|A) \times P(A))}{P(B)} \quad \boxed{[\text{Bayes' Theorem}]}$$

Are Male viewers and AOT independent?

$$P(M \cap AOT) = 0.16$$

$$P(M) = 0.46$$

$$P(AOT) = 0.40$$

$$P(A \cap B) = P(A) \times P(B)$$

$$P(M) \times P(AOT) = 0.46 \times 0.40 = 0.184$$

Since $P(M \cap AOT) \neq P(M) \times P(AOT)$, so not independent.

	Smart		Not Smart	
	Study	Not Study	Study	Not Study
Prepared	0.432	0.16	0.084	0.008
Not Prepared	0.048	0.16	0.036	0.072

Is Smart conditionally independent of prepared given study?

$$P(SM \cap P_{RE} | STU) = \frac{P(SM \cap P_{RE} \cap STU)}{P(STU)} = \frac{0.432}{0.6} = 0.72$$

Again,

$$P(SM \wedge PnE | STU) = P(SM | STU) * P(PnE | STU) = (0.8)^2 = 0.64$$

$$P(SM | STU) = \frac{P(SM \wedge STU)}{P(STU)} = \frac{0.432}{0.6} = 0.72$$

$$P(PnE | STU) = \frac{P(PnE \wedge STU)}{P(STU)} = \frac{0.516}{0.6} = 0.86$$

$$\therefore P(SM | STU) * P(PnE | STU) = 0.8 \times 0.86 = 0.68 \neq 0.72$$

\therefore Not conditionally independent

$$J1.0 = (T00 \wedge M)^2$$

$$J2.0 = (M)^2$$

A person is brought in front of a jury. The jury

finds the defendant guilty in 98% of the cases in which he committed a crime and it finds the defendant not guilty 97% of the cases when the defendant has not committed a crime. Only 0.8% of the population has committed a crime.

If a random person is found guilty by the jury what is more likely, criminal or not?

Subject: Statistics & Probability Date: 26/01/2023

$$P(J1.0) = \frac{(J1.0 \wedge T00 \wedge M)^2}{(J1.0)^2} = (J1.0 | T00 \wedge M)^2$$

Bayes' TheoremNaive Bayes

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

prior belief of $A \leftarrow$ and A prior belief of $B \leftarrow$ and B prior belief of $A \leftarrow$ and B Problem 1:

Bag A contains 3 red balls
4 black balls

Bag B contains 5 red balls
6 black balls

One ball is drawn at random from one of the bags, and it is red.
Find the probability that it was from bag B.

Solution:

Bag A: (3 Red) \times (4 Black) \times (A) $\frac{1}{2}$

Bag B: 5 Red and 6 Black

$P(\text{Bag B given that ball is red}) = P(B_B | \text{Red})$

$$P(B_A) = \frac{1}{2}, P(B_B) = \frac{1}{2}$$

$$P(\text{Red} | B_A) = \frac{3}{7}, P(\text{Red} | B_B) = \frac{5}{11}$$

$$\begin{aligned} &= \frac{P(B_B) \cdot P(\text{Red} | B_B)}{P(B_A) \cdot P(\text{Red} | B_A) + P(B_B) \cdot P(\text{Red} | B_B)} \\ &= \frac{\frac{1}{2} \times \frac{5}{11}}{\left(\frac{1}{2} \times \frac{3}{7}\right) + \left(\frac{1}{2} \times \frac{5}{11}\right)} \\ &= \frac{25}{68} \end{aligned}$$

Problem 2:

Given three identical boxes A, B, C.

A has \rightarrow 2 gold coins

B has \rightarrow 2 silver coins

C has \rightarrow 1 gold and 1 silver coin.

$$(A)^2 + (A/B)^2 = (A/A)^2$$

$$(A)^2$$

What is the probability that the other coin in the box is also of gold if a person choose a box at random and take out a coin which is gold?

$$P(B) = \frac{1}{3}, P(A) = \frac{1}{3}, P(C) = \frac{1}{3}$$

$$P(A|B) = \frac{0}{2}, P(A|A) = \frac{2}{2}, P(A|C) = \frac{1}{2}$$

$$P(A|C) = \frac{P(A) \times P(A|A)}{P(A) \times P(A|A) + P(B) \times P(A|B) + P(C) \times P(A|C)}$$

$$= \frac{\frac{1}{3} \times \frac{2}{2}}{\frac{1}{3} \times \frac{2}{2} + \frac{1}{3} \times 0 + \frac{1}{3} \times \frac{1}{2}}$$

$$= \frac{\frac{1}{3} \times \frac{2}{2}}{\left(\frac{1}{3} \times \frac{2}{2}\right) + \left(\frac{1}{3} \times 0\right) + \left(\frac{1}{3} \times \frac{1}{2}\right)}$$

$$= \frac{\frac{1}{3} \times \frac{2}{2}}{\left(\frac{1}{3} \times \frac{2}{2}\right) + \left(\frac{1}{3} \times 0\right) + \left(\frac{1}{3} \times \frac{1}{2}\right)}$$

$$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$= \frac{\frac{1}{3} \times \frac{2}{2}}{\left(\frac{1}{3} \times \frac{2}{2}\right) + \left(\frac{1}{3} \times 0\right) + \left(\frac{1}{3} \times \frac{1}{2}\right)}$$

$$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

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

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

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

Problem 3:

Bag X → contains 1 White, 2 Red, 3 Green balls

Bag Y → contains 2 White, 3 Red, 1 Green balls

Bag Z → contains 3 White, 1 Red, 2 Green balls

Two balls were chosen from the bag and it was W and R. Find the probability that the balls so drawn came from the Bag Y.

Let E = getting 1W and 1R

$$\begin{aligned}
 P(BY|E) &= \frac{P(Y) \times P(E|Y)}{P(X) \times P(E|X) + P(Y) \times P(E|Y) + P(Z) \times P(E|Z)} \\
 &= \frac{\frac{1}{3} \times \left(2C_1 \times 3C_1 \right) / 6C_2}{\frac{1}{3} \times \left(1C_1 \times 2C_1 \right) / 6C_2 + \frac{1}{3} \times \left(2C_1 \times 3C_1 \right) / 6C_2 + \frac{1}{3} \times \left(3C_1 \times 1C_1 \right) / 6C_2} \\
 &= \frac{6}{11}
 \end{aligned}$$

Problem 4:

Suppose that 5% of men and 0.25% of women have grey hair. A grey haired person selected at random. What is the probability of this person being male?

There are equal numbers of males and females.

$$P(A|M) = \frac{5/100}{100/100} \quad P(A|W) = \frac{0.25/100}{100/100}$$

$$P(M|A) = \frac{P(A|M) \cdot P(M)}{P(A|M) \cdot P(M) + P(A|W) \cdot P(W)}$$

$$= \frac{\frac{5/100}{100} \times \frac{1}{2}}{\left(\frac{5/100}{100} \times \frac{1}{2}\right) + \left(\frac{0.25/100}{100} \times \frac{1}{2}\right)}$$

$$= \frac{1/40}{1/40 + 1/800}$$

$$= \frac{20}{21}$$

Problem 5:

In a certain college, 4% of the boys and 1% of girls are taller than 1.8m. 60% of the students are girls. If a student is selected at random and is found to be taller than 1.8m. What is the probability that the student is a girl?

E = taller than 1.8m

$$P(E/B) = \frac{4}{100}, P(E/a) = \frac{1}{100}, P(B) = \frac{40}{100}, P(a) = \frac{60}{100}$$

$$P(a/E) = \frac{P(E|a) \cdot P(a)}{P(E|a) \cdot P(a) + P(E|B) \cdot P(B)}$$

$$= \frac{\frac{1}{100} \times \frac{60}{100}}{\left(\frac{1}{100} \times \frac{60}{100}\right) + \left(\frac{4}{100} \times \frac{40}{100}\right)} = \frac{60}{220} = 0.2727 = 27.27\%$$

$$P(B) \left(\frac{0.04}{0.01} \times \frac{0.01}{0.01} \right) = (13|A) 9 \quad \left(\frac{0.08}{0.01} \times \frac{0.01}{0.01} \right) = (5|A) 9$$

$$13.0 = 18.0 =$$

$$\frac{8.0 \times 18.0}{(8.0 \times 18.0) + (8.0 \times 18.0)} = \frac{144}{288} = (A|a) 9$$

$$28.0 = \frac{18}{28}$$

Problem 6:

If a machine is correctly setup, it produces 90% acceptable items. If it is incorrectly setup, it produces only 40% acceptable items. Past experience shows that 80% of the setups are correctly done. After a setup the machine produces 2 acceptable items.

Find the probability that the machine is correctly setup.

$A = \text{getting 2 acceptable items}$

$$P(C|A) = \frac{P(A|C) \times P(C)}{P(A|C) \times P(C) + P(A|C') \times P(C')}$$

$$P(C) = \frac{80}{100} \quad P(C') = \left(1 - \frac{80}{100}\right) \\ = 0.8 \quad = 0.2$$

$$P(A|C) = \left(\frac{90}{100} \times \frac{90}{100}\right) \quad P(A|C') = \left(\frac{40}{100} \times \frac{40}{100}\right) \\ = 0.81 \quad = 0.16$$

$$P(C|A) = \frac{0.81 \times 0.8}{(0.81 \times 0.8) + (0.16 \times 0.2)} \\ = \frac{81}{85} = 0.95$$

Problem 7:

: 9 marks

In a factory which manufactures bolts, machine A, B, C manufactures respectively 25%, 35%, 40% of the bolts. The output 5%, 4%, 2% are defective bolts. A bolt is drawn at random and found to be defective. What is the probability that it is manufactured by the machine B?

$$P(B|D) = \frac{P(D|B) \times P(B)}{P(D|B) \times P(B) + P(D|A) \times P(A) + P(D|C) \times P(C)}$$

$$P(A) = \frac{1}{3}, P(B) = \frac{1}{3}, P(C) = \frac{1}{2}$$

$$P(D|A) = \frac{25}{100} \times \frac{1}{100} = \frac{25}{100} \times \frac{1}{100} = \frac{40}{100} \times \frac{1}{100} \times \frac{1}{100} =$$

$$P(D|B) = \frac{5}{100}, P(D|C) = \frac{2}{100}$$

$$P(B|D) = \frac{\left(\frac{1}{100} \times \frac{35}{100}\right)}{\left(\frac{1}{100} \times \frac{35}{100}\right) + \left(\frac{5}{100} \times \frac{25}{100}\right) + \left(\frac{2}{100} \times \frac{40}{100}\right)}$$

$$= 0.04 \times 0.35$$

$$= \frac{(0.04 \times 0.35) + (0.05 \times 0.25) + (0.02 \times 0.4)}{(0.04 \times 0.35) + (0.05 \times 0.25) + (0.02 \times 0.4)}$$

$$= 0.405797 + 0.1875 + 0.08 = 0.673297$$

Problem 8:

Companies B1, B2, B3 produces 30%, 45%, 25% of the cars respectively. It is known that 2%, 3% and 2% of these cars produced from are defective.

(a) What is the probability that a car purchased is defective?

$$P(B_1) = \frac{30}{100}, \quad P(B_2) = \frac{45}{100}, \quad P(B_3) = \frac{25}{100}$$

$$P(D|B_1) = \frac{2}{100}, \quad P(D|B_2) = \frac{3}{100}, \quad P(D|B_3) = \frac{2}{100}$$

$$P(D) = P(D|B_1) P(B_1) + P(D|B_2) P(B_2) + P(D|B_3) P(B_3)$$

$$= \left(\frac{2}{100} \times \frac{30}{100} \right) + \left(\frac{3}{100} \times \frac{45}{100} \right) + \left(\frac{2}{100} \times \frac{25}{100} \right)$$

$$= (0.02 \times 0.3) + (0.03 \times 0.45) + (0.02 \times 0.25)$$

$$= 0.0245$$

(b) If a car purchased is found out to be defective, what is the probability that this car is produced by Company B1?

$$P(B_1|D) = \frac{P(D|B_1) \times P(B_1)}{P(D|B_1) \times P(B_1) + P(D|B_2) \times P(B_2) + P(D|B_3) \times P(B_3)}$$

$$P(B_1) = \frac{30}{100}, P(B_2) = \frac{45}{100}, P(B_3) = \frac{25}{100} = (111)^9$$

$$P(D|B_1) = \frac{2}{100}, P(D|B_2) = \frac{3}{100}, P(D|B_3) = \frac{4}{100}$$

$$P(B_1|D) = \frac{\frac{2}{100} \times \frac{30}{100}}{\left(\frac{2}{100} \times \frac{30}{100}\right) + \left(\frac{3}{100} \times \frac{45}{100}\right) + \left(\frac{4}{100} \times \frac{25}{100}\right)}$$

$$= \frac{0.02 \times 0.3}{(0.02 \times 0.3) + (0.03 \times 0.45) + (0.02 \times 0.25)}$$

$$= \frac{6 \times 10^{-3}}{6 \times 10^{-3} + 0.0135 + 5 \times 10^{-3}}$$

$$= \frac{6 \times 10^{-3}}{0.0245}$$

$$= 0.244897$$

Problem: 9

A lot of IC chips contains 2% defective chips. Each is tested before delivery. The tester itself is not totally reliable. Probability of tester says the chip is good when it is really good is 0.95 and the probability of tester says a chip is defective when it is actually defective is 0.99. If a tested device is indicated to be defective, what is the probability that it is actually defective.

$$P(T|D) = \frac{P(D|T) \times P(T)}{P(D|T) \times P(T) + P(D|T^c) \times P(T^c)}$$

~~$\frac{2/100}{2/100 + 0.95 \times 0.02} = \frac{2/100}{2/100 + 0.019} = \frac{2/100}{2.019} = \frac{2}{2.019} = 0.995$~~

$T = \text{tested say good}$
 $T^c = \text{untested say defective}$

$$P(D) = \frac{2}{100}, P(D^c) = 1 - \frac{2}{100}$$

$$P(D|T^c) = 0.95, P(TT|D) = 0.94, P(TT|D^c) = 1 - P(T|D^c)$$

$$= 1 - 0.95$$

$$= 0.05$$

$$P(D|T^c) = \frac{P(TT|D) \times P(D)}{P(TT|D) \times P(D) + P(TT|D^c) \times P(D^c)}$$

$$= \frac{0.94 \times 0.02}{(0.94 \times 0.02) + (0.05 \times 0.98)}$$

$$= \frac{0.0188}{0.0188 + 0.049} = \frac{0.0188}{0.0678} = 0.27$$

Q: What are

- 1) The 3 types of variables that are measured in the test A
- 2) What is the relationship between the variables A and B provided in the table
- 3) If we have been assigned with 300 subjects to utilize a new technique A and 200 subjects to utilize a new technique B, then what is the probability that the new technique A will be better than the new technique B
- 4) If we have 100 subjects in the study, then what is the probability that the new technique A will be better than the new technique B
- 5) If we have 100 subjects in the study, then what is the probability that the new technique A will be better than the new technique B
- 6) If we have 100 subjects in the study, then what is the probability that the new technique A will be better than the new technique B
- 7) If we have 100 subjects in the study, then what is the probability that the new technique A will be better than the new technique B
- 8) If we have 100 subjects in the study, then what is the probability that the new technique A will be better than the new technique B
- 9) If we have 100 subjects in the study, then what is the probability that the new technique A will be better than the new technique B
- 10) If we have 100 subjects in the study, then what is the probability that the new technique A will be better than the new technique B

(P(A) = 6/14 (Naive Bayes) \rightarrow P(A) = 9/14 (Naive Bayes))

Dataset

Day	Outlook	Temperature	Humidity	Windy	Play Tennis
1	Sunny	Hot	H	W	No
2	Sunny	Hot	H	S	No
3	Overcast	Hot	H	W	Yes
4	Rain	Mild	H	W	Yes
5	Rain	Cool	N	W	Yes
6	Rain	Cool	N	S	No
7	Overcast	Cool	N	S	Yes
8	Sunny	Mild	H	W	No
9	Sunny	Cool	N	W	Yes
10	Rain	Mild	N	W	Yes
11	Sunny	Mild	N	S	Yes
12	Overcast	Mild	H	S	Yes
13	Overcast	Hot	N	W	Yes
14	Rain	Mild	N	S	No

$$P(\text{Play Tennis} = \text{Yes}) = \frac{9}{14} = 0.64$$

$$P(\text{Play Tennis} = \text{No}) = \frac{5}{14} = 0.36$$

Outlook	Y	N
Sunny	2/9	3/5
Overcast	4/9	0/14
Rain	3/9	2/5

Humid	Y	N
H	3/9	4/5
N	6/9	1/5

Windy	Y	N
W	6/9	2/5
S	3/9	3/5

Temp	Y	N
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	3/5

(Outlook = sunny, Temp = cool, Hum = high, Wind = strong)

$$V_{NB} = \underset{v_j \in \{yes, No\}}{\operatorname{argmax}} P(v_j) \prod_{i=1}^4 P(a_i | v_j)$$

$$= \underset{v_j \in \{yes, No\}}{\operatorname{argmax}} P(v_j)$$

$V_{NB}(yes)$:

$$P(yes) = P(sunny | y) P(cool | y) P(high | y) P(strong | y)$$

$$= 0.64 \times 0.222 \times 0.3333 \times 0.3333 \times 0.3333$$

$$= 0.00526$$

$V_{NB}(No)$:

$$P(No) = P(sunny | N) P(cool | N) P(high | N) P(strong | N)$$

$$= 0.36 \times 0.6 \times 0.2 \times 0.8 \times 0.6$$

$$= 0.020736$$

$$P(d.o) = \frac{P(y)}{P(y)} = (0.00526 + 0.020736) = 0.025996$$

$$V_{NB}(Yes) = \frac{V_{NB}(yes)}{V_{NB}(yes) + V_{NB}(No)} = \frac{0.00526}{0.00526 + 0.020736} = 0.00526 / 0.025996 = 0.02023$$

$$V_{NB}(No) = \frac{V_{NB}(No)}{V_{NB}(Yes) + V_{NB}(No)} = \frac{0.020736}{0.00526 + 0.020736} = 0.020736 / 0.025996 = 0.79766$$

Dataset

No	Color	Legs	Heights	Smelly	Species
1	White	3	S	Y	M
2	Green	2	LT	N	M
3	Green	3	LS	Y	M
4	White	3	LS	Y	M
5	Green	2	LS	N	H
6	White	2	LT	N	H
7	White	2	LT	N	H
8	White	2	S	Y	H

$$P(M) = \frac{4}{8} = 0.5$$

$$P(H) = \frac{4}{8} = 0.5$$

Color	M	H
W	2/4	3/4
G	2/4	1/4

Legs	M	H
2	1/4	4/4
3	3/4	0/4

Height	M	H
T	3/4	3/4
S	1/4	3/4

Smell	M	H
Y	3/4	1/4
S	1/4	3/4

(Color = Green ; Legs = 2, Height = Tall, Smelly = No)

$$P(M) = P(M) \times P(G|M) \times P(2|G) \times P(Tall|M) \times P(No|G)$$

$$= 0.5 \times \frac{1}{2} \times \frac{1}{4} \times \frac{3}{4} \times \frac{1}{2} = 0.0117$$

$$P(H) = P(H) \times P(G|H) \times P(2|H) \times P(Tall|H) \times P(No|H)$$

$$= 0.5 \times \frac{1}{2} \times \frac{1}{4} \times \frac{3}{4} \times \frac{1}{2} = 0.047$$

Decision Tree

Day	Outlook	Temp	Humidity	Wind	Play Tennis
01	Sunny	Hot	High	Weak	No
02	Sunny	Hot	High	Strong	No
03	Overcast	Hot	High	Weak	Yes
04	Rain	Mild	High	Weak	Yes
05	Rain	Cold	Normal	Weak	Yes
06	Rain	Cold	Normal	Strong	No
07	Overcast	Cold	Normal	Strong	Yes
08	Sunny	Mild	High	Weak	No
09	Sunny	Cold	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

$$\text{Entropy} = - \sum_{i=1}^n P_i \log_2 P_i$$

$$E(\text{Decision}) = -P(\text{Yes}) \log_2 P(\text{Yes}) + (-P(\text{No}) \log_2 P(\text{No}))$$

$$= \left(\frac{9}{14} \log_2 \left(\frac{9}{14} \right) + \left(-\frac{5}{14} \log_2 \left(\frac{5}{14} \right) \right) \right) = 0.9402859$$

$$= 0.9402859 = P_1 \times P_2 \times P_3 \times P_4 \times 2.0 =$$

$$(H \mid M) = (H \mid \text{Outlook}) = (H \mid T) = (H \mid S) = (H \mid O) = (H)$$

$$= P_1 \times P_2 \times P_3 \times P_4 \times 2.0 =$$

$E(\text{Yes|Sunny}) \Rightarrow$

[outlook]

$$\begin{aligned}
 & -P(\text{Yes|Sunny}) \log_2 (\text{Yes|Sunny}) - P(\text{No|Sunny}) \log_2 (\text{No|Sunny}) \\
 & = -\frac{2}{5} \log_2 \left(\frac{2}{5}\right) - \frac{3}{5} \log_2 \left(\frac{3}{5}\right) - \left(\frac{3}{5}\right) \log_2 \frac{3}{5} \\
 & = \approx 0.971
 \end{aligned}$$

$E(\text{Overcast}) \Rightarrow$

$$\begin{aligned}
 & -P(\text{Yes|Overcast}) \log_2 (\text{Yes|Overcast}) - P(\text{No|Overcast}) \log_2 (\text{No|Overcast}) \\
 & = -\frac{4}{4} \log_2 \left(\frac{4}{4}\right) - \frac{0}{4} \log_2 \left(\frac{0}{4}\right) - \left(\frac{4}{4}\right) \log_2 \frac{4}{4} \\
 & = 0
 \end{aligned}$$

$E(\text{Rain}) \Rightarrow$

$$\begin{aligned}
 & -P(\text{Yes|Rain}) \log_2 (\text{Yes|Rain}) - P(\text{No|Rain}) \log_2 (\text{No|Rain}) \\
 & = -\frac{3}{5} \log_2 \left(\frac{3}{5}\right) - \frac{2}{5} \log_2 \left(\frac{2}{5}\right) - \left(\frac{3}{5}\right) \log_2 \frac{3}{5} \\
 & = 0.971
 \end{aligned}$$

Information Gain: [outlook]

$$\begin{aligned}
 & E(\text{Decision}) - P(\text{Sunny}) \times E(\text{Sunny}) - P(\text{Rain}) \times E(\text{Rain}) - P(\text{Oc}) \times E(\text{Oc}) \\
 & = 0.9402 - \left(\frac{5}{14} \times 0.971\right) - \left(\frac{5}{14} \times 0.971\right) - \left(\frac{4}{14} \times 0\right) \\
 & = 0.246
 \end{aligned}$$

$E(\text{Hot}) \Rightarrow$

[Temp]

 $\Leftarrow (\text{Normal Hot}) \exists$

$$-P(\text{Yes}|\text{Hot}) \log_2 (\text{Yes}|\text{Hot}) - P(\text{No}|\text{Hot}) \log_2 (\text{No}|\text{Hot})$$

$$= -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right)$$

$$= 1.0$$

I.F.C. 0

 $E(\text{Mild}) \Rightarrow$ $\Leftarrow (\text{Normal Mild}) \exists$

$$-P(\text{Yes}|\text{Mild}) \log_2 (\text{Yes}|\text{Mild}) - P(\text{No}|\text{Mild}) \log_2 (\text{No}|\text{Mild})$$

$$= -\frac{4}{6} \log_2 \left(\frac{4}{6}\right) - \frac{2}{6} \log_2 \left(\frac{2}{6}\right)$$

$$= -\frac{4}{6} \log_2 \left(\frac{4}{6}\right) - \frac{2}{6} \log_2 \left(\frac{2}{6}\right)$$

$$= 0.918295$$

 $\Leftarrow (\text{Mild}) \exists$ $E(\text{Cold}) \Rightarrow$

$$-P(\text{Yes}|\text{Cold}) \log_2 (\text{Yes}|\text{Cold}) - P(\text{No}|\text{Cold}) \log_2 (\text{No}|\text{Cold})$$

$$= -\frac{2}{4} \log_2 \left(\frac{2}{4}\right) - \left(\frac{1}{4}\right) \log_2 \left(\frac{1}{4}\right)$$

$$= 0.811278$$

 $\Leftarrow (\text{Cold}) \exists$

Information gain :

$$E(D) = P(\text{Hot}) \times E(\text{Hot}) + P(\text{Mild}) \times E(\text{Mild}) + P(\text{Cold}) \times E(\text{Cold})$$

$$= 0.9402 - \left(\frac{4}{14} \times 1\right) - \left(\frac{6}{14} \times 0.918\right) - \left(\frac{4}{14} \times 0.811278\right) = 0.0289$$

[Humidity]

 $E(High) \Rightarrow$

$$-P(Yes|High) \log_2 (Yes|High) - P(No|High) \log_2 (No|High)$$

$$= -\frac{3}{7} \log_2 \left(\frac{3}{7}\right) - \left(\frac{4}{7}\right) \log_2 \left(\frac{4}{7}\right) = \left(\frac{3}{8}\right)_{S.P.} \frac{3}{8} =$$

$$= 0.985228136$$

 $E(Normal) \Rightarrow$

$$-P(Yes|Normal) \log_2 (Yes|Normal) - P(No|Normal) \log_2 (No|Normal)$$

$$= -\frac{6}{7} \log_2 \left(\frac{6}{7}\right) - \frac{1}{7} \log_2 \left(\frac{1}{7}\right) = \left(\frac{6}{8}\right)_{S.P.} \frac{6}{8} =$$

$$= 0.59167277$$

Information Gain:

: ~~using~~ no information

$$E(D) = P(High) \times E(High) + P(Normal) \times E(Normal)$$

$$= 0.94028 - \left(\frac{7}{14} \times 0.9852\right) - \left(\frac{7}{14} \times 0.59167\right)$$

$$= 0.1616$$

[Wind]

$E(\text{Weak}) \Rightarrow$

$\Leftarrow (A \cap B) \cap$

$$\begin{aligned}
 & -P(\text{Yes} \mid \text{Weak}) \log_2 (\text{Yes} \mid \text{Weak}) - P(\text{No} \mid \text{Weak}) \log_2 (\text{No} \mid \text{Weak}) \\
 & = -\frac{6}{8} \log_2 \left(\frac{6}{8}\right) - \frac{2}{8} \log_2 \left(\frac{2}{8}\right) = \left(-\frac{6}{8}\right) \log_2 \frac{6}{8} + \frac{2}{8} \log_2 \frac{2}{8} \\
 & = 0.8112781
 \end{aligned}$$

261888280.0

$E(\text{Strong}) \Rightarrow$

$\Leftarrow (A \cap B) \cap$

$$\begin{aligned}
 & -P(\text{Yes} \mid \text{Strong}) \log_2 (\text{Yes} \mid \text{Strong}) - P(\text{No} \mid \text{Strong}) \log_2 (\text{No} \mid \text{Strong}) \\
 & = -\frac{3}{6} \log_2 \left(\frac{3}{6}\right) - \left(-\frac{3}{6}\right) \log_2 \left(\frac{3}{6}\right) = \left(-\frac{3}{6}\right) \log_2 \frac{3}{6} + \left(-\frac{3}{6}\right) \log_2 \frac{3}{6} \\
 & = 1.0
 \end{aligned}$$

PPSF2182.0

Information Gain:

$$\begin{aligned}
 E(I) & \leftarrow P(\text{Weak}) \times E(\text{Weak}) + P(\text{Strong}) \times E(\text{Strong}) \\
 & = 0.9402859 - \left(\frac{8}{14} \times 0.8112781\right) - \left(\frac{6}{14} \times 1\right) \\
 & = 0.0478
 \end{aligned}$$

Summary:

$$Gain(\text{Outlook}) = 0.246$$

$$Gain(\text{Temp}) = 0.0289$$

$$Gain(\text{Humidity}) = 0.1516$$

$$Gain(\text{Wind}) = 0.0478$$

Outlook:

Sunny: 01, 02, 08, 09, 11 $\rightarrow X$

Overcast: 03, 07, 12, 13 $\rightarrow \text{G}$

Rain: 04, 05, 06, 10, 14 $\rightarrow R$

Outlook = sunny

Day	Temp	Humidity	Wind	Play Tennis
D1	Hot	High	Weak	No
D2	Hot	High	Strong	No
D8	Mild	High	Weak	No
D9	Cool	Normal	Weak	Yes
D11	Mild	Normal	Strong	Yes

Temp

$$\begin{aligned}
 E(\text{Sunny}) &= -P(\text{Yes}) \log_2 P(\text{Yes}) - P(\text{No}) \log_2 P(\text{No}) \\
 &= -\frac{2}{5} \log_2 \left(\frac{2}{5}\right) - \frac{3}{5} \log_2 \left(\frac{3}{5}\right) \\
 &= 0.97
 \end{aligned}$$

$$\begin{aligned}
 E(\text{Hot}) &= -P(\text{Yes}|\text{Hot}) \log_2 P(\text{Yes}|\text{Hot}) - P(\text{No}|\text{Hot}) \log_2 P(\text{No}|\text{Hot}) \\
 &= -\frac{0}{2} \log_2 \left(\frac{0}{2}\right) - P\left(\frac{2}{2}\right) \log_2 \left(\frac{2}{2}\right) \\
 &= 0
 \end{aligned}$$

$$E(\text{Mild}) = -P(\text{Yes}|\text{Mild}) \log_2 (\text{Yes}|\text{Mild}) - P(\text{No}|\text{Mild}) \log_2 (\text{No}|\text{Mild})$$

$$= -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right)$$

$$= 1.0$$

$$E(\text{Cold}) = -P(\text{Yes}|\text{Cold}) \log_2 (\text{Yes}|\text{Cold}) - P(\text{No}|\text{Cold}) \log_2 (\text{No}|\text{Cold})$$

$$= -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) \log_2 \left(\frac{1}{2}\right)$$

$$= 0$$

Information Gain:

$$E(S) = P(\text{Hot}) \times E(\text{Hot}) + P(\text{Mild}) \times E(\text{Mild}) + P(\text{Cold}) \times E(\text{Cold})$$

$$= 0.97 - 0 - 1 \times \frac{2}{5} - 0$$

$$= 0.570$$

Humidity

$$E(\text{High}) = -P(\text{Yes}|\text{High}) \log_2 (\text{Yes}|\text{High}) - P(\text{No}|\text{High}) \log_2 (\text{No}|\text{High})$$

$$= -\frac{0}{3} \log_2 \left(\frac{0}{3}\right) - \left(\frac{3}{3}\right) \log_2 \left(\frac{3}{3}\right)$$

$$= 0$$

$$E(\text{Normal}) = -P(\text{Yes}|\text{Normal}) \log_2 (\text{Yes}|\text{Normal}) - P(\text{No}|\text{Normal}) \log_2 (\text{No}|\text{Normal})$$

$$= -\left(\frac{1}{2}\right) \log_2 \left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) \log_2 \left(\frac{1}{2}\right)$$

$$= 0$$

So:

$$E(S) = P(H) \times E(H) + P(N) \times E(N) = 0.97 - 0 - 0$$

$$= 0.97$$

$$\begin{aligned}
 E(\text{Weak}) &= -P(\text{Yes}|\text{Weak}) \log_2 (\text{Yes}|\text{Weak}) - P(\text{No}|\text{Weak}) \log_2 (\text{No}|\text{Weak}) \\
 &= -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) \\
 &= 0.9182
 \end{aligned}$$

$$\begin{aligned}
 E(\text{Strong}) &= -P(\text{Yes}|\text{Strong}) \log_2 (\text{Yes}|\text{Strong}) - P(\text{No}|\text{Strong}) \log_2 (\text{No}|\text{Strong}) \\
 &= -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right) \\
 &= 1.0
 \end{aligned}$$

Information Gain:

Wind

$$E(S) - P(W) \times E(W) - P(S) \times E(S)$$

$$0.971 - \frac{3}{5} \times 0.9182 - \frac{2}{5} \times 1$$

$$\Rightarrow 0.02008$$

Outlook = Rain

Summary:

$$\text{Gain(Temp)}: 0.570$$

$$\text{Gain(Humidity)}: 0.97$$

$$\text{Gain(Wind)}: 0.02008$$

$$E(\text{Hot}) = 0$$

$$\begin{aligned}
 E(\text{Mild}) &= -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) \\
 &= 0.91829
 \end{aligned}$$

$$\begin{aligned}
 E(\text{Cold}) &= -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right) \\
 &= 1.0
 \end{aligned}$$

$$E(\text{Hot}) = 0.0$$

Day	Temp	Humidity	Wind	Play Tennis
D4	Mild	High	Weak	Yes
D5	Cold	Normal	Weak	Yes
D6	Cold	Normal	Strong	No
D10	Mild	Normal	Weak	Yes
D14	Mild	High	Strong	No

$$\begin{aligned}
 E(\text{Rain}) &= -\frac{3}{5} \log_2 \left(\frac{3}{5}\right) - \frac{2}{5} \log_2 \left(\frac{2}{5}\right) \\
 &= 0.970950
 \end{aligned}$$

$$E(\text{High}) = -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{2} \log_2 \left(\frac{1}{2}\right) = 1.0$$

$$E(\text{Normal}) = -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) = 0.9182$$

Information Gain: [Temp]

$$\begin{aligned}
 E(\text{Rain}) &= P(\text{Hot}) \times E(\text{Hot}) + P(\text{Mild}) \times E(\text{Mild}) + P(\text{Cold}) \times E(\text{Cold}) \\
 &= 0.970 \times 0 + \left(\frac{3}{5} \times 0.918\right) + \left(\frac{2}{5} \times 1.0\right) \\
 &= 0.0192
 \end{aligned}$$

Information Gain: [Humidity]

$$\begin{aligned}
 E(\text{Rain}) &= P(\text{High}) \times E(\text{High}) + P(\text{Normal}) \times E(\text{Normal}) \\
 &= 0.970 \times \left(\frac{2}{5} \times 1.0\right) + \left(\frac{3}{5} \times 0.918\right) \\
 &= 0.0192
 \end{aligned}$$

$$E(\text{Weak}) = -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \left(\frac{1}{3} \log_2 \left(\frac{1}{3}\right)\right) = 0.0$$

$$E(\text{Strong}) = -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \left(\frac{1}{2} \log_2 \left(\frac{1}{2}\right)\right) = 0.0$$

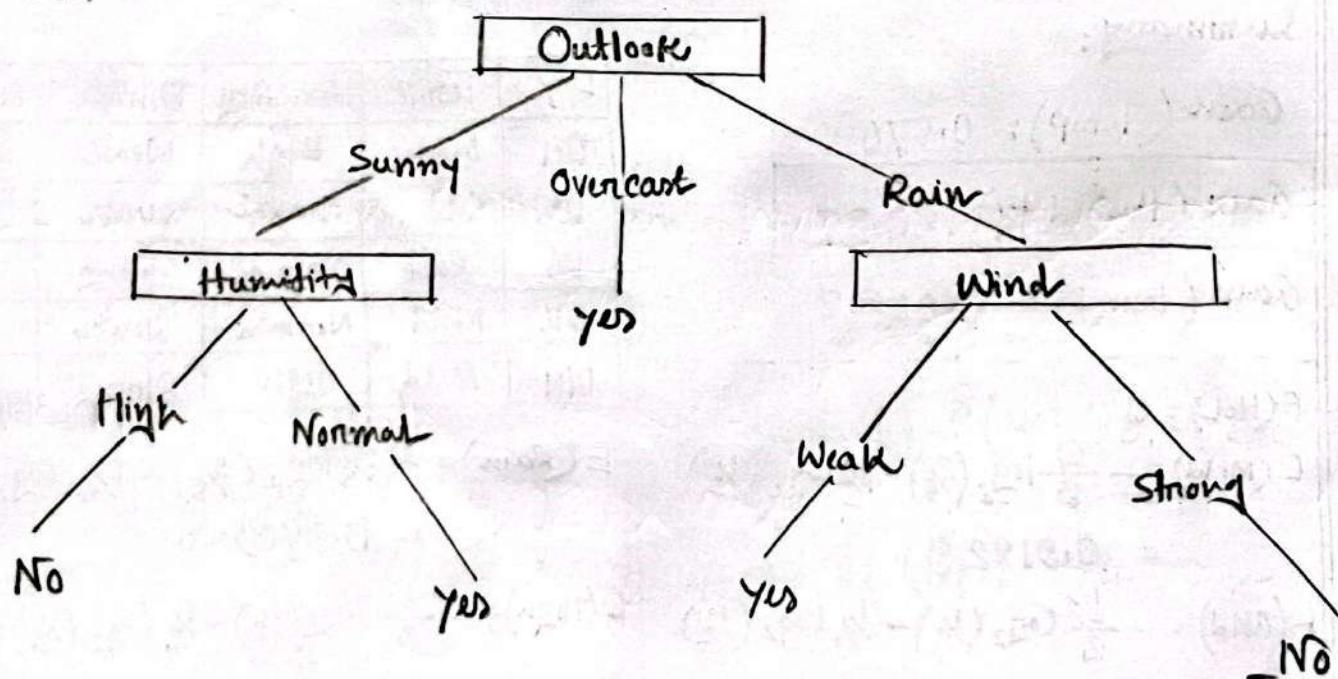
$$\text{Temp} = 0.0192$$

$$\text{Humidity} = 0.0192$$

$$\text{Wind} = 0.02$$

Information Gain: [Wind]

$$\begin{aligned}
 E(\text{Rain}) &= P(\text{Weak}) \times E(\text{Weak}) + P(\text{Strong}) \times E(\text{Strong}) \\
 &= 0.97 \times 0 + 0 \\
 &= 0.07
 \end{aligned}$$



Problems

①

A patient went to the hospital for a malaria test. The doctor informed him that their test can successfully diagnose malaria positive given that patient is actually malaria positive 94% of the time.

Also, the probability of having no malaria and getting a malaria negative test result is 4%.

Meanwhile, 27% of people in general who come for tests are malaria positive. Now if the patient is already diagnosed malaria negative, then calculate the probability of the patient actually being malaria negative.

Given that,

$$P(\text{Positive test} \mid \text{Actually positive}) = 0.94$$

$$P(\text{Negative test} \mid \text{Actually negative}) = 0.04$$

$$P(\text{Actually positive}) = 0.27$$

$$P(\text{Actually negative}) = 1 - 0.27 = 0.73$$

$$P(\text{Actually negative} \mid \text{Negative test}) =$$

$$= \frac{P(\text{Negative test} \mid \text{Actually negative}) \times P(\text{Actually negative})}{P(\text{Negative test})}$$

$$P(\text{Negative test}) = P(\text{Negative test} \mid \text{Actually negative}) \times P(\text{Actually negative}) + P(\text{Negative test} \mid \text{Actually positive}) \times P(\text{Actually positive})$$

$$P(\text{Negative test} \mid \text{Actually positive}) = 1 - 0.04 = 0.96$$

$$P(\text{Actually negative} \mid \text{Negative test}) = \frac{(0.04 \times 0.73)}{(0.04 \times 0.73) + (0.06 \times 0.27)}$$

$$= \frac{0.0292}{0.0292 + 0.0162}$$

$$= 0.64317$$

(b) Is being malaria positive and having a positive test result independent of each other?

$$\text{We know, } P(A \cap B) = P(A) \times P(B)$$

$$= 0.27 \times 0.94$$

Assume,
 $A = \text{Malaria positive}$

$B = \text{Malaria positive test result}$

$$\text{Given, } P(A) = 0.27$$

$$\begin{aligned}
 P(B) &= P(\text{positive test}) \\
 &= P(B|A) \times P(A) + P(B|A') \times P(A') \\
 &= P(\text{positive test} | \text{Actually positive}) \times P(\text{Actually positive}) \\
 &\quad + P(\text{positive test} | \text{Actually negative}) \times P(\text{Actually negative}) \\
 &= (0.94 \times 0.27) + ([1 - 0.04] \times 0.73) \\
 &= 0.2538 + 0.4538 \times 0.73 \\
 &= 0.2976 \quad 0.9546
 \end{aligned}$$

$$P(A \cap B) = 0.2538$$

$$P(A) = 0.27$$

$$P(B) = 0.9546$$

$$\begin{aligned}
 P(A) \times P(B) &= 0.27 \times 0.9546 \\
 &= 0.257742
 \end{aligned}$$

$$P(A \cap B) \neq P(A) \times P(B)$$

They are not independent.

$$\frac{0.2538}{(\text{Not visiting})^2} = 28.0 = \frac{0.27 \times 0.9546}{(\text{Not visiting})^2}$$

Covid-19 test all over the World aren't 100% accurate.

A patient is actually positive in 85% of the cases

When the test comes out to be positive. A person is actually positive in 10% of the cases. When the test comes out to be negative.

- ① Of all the people who tested for Covid-19, 70% of them actually had the disease. If 1000 people participated in the tests, calculate the probability of a person's test results being positive.

$$P(\text{Actually positive} \mid \text{positive test}) = 0.85$$

$$P(\text{Actually positive} \mid \text{negative test}) = 0.10$$

$$P(\text{Actually positive}) = 0.70$$

$$\text{Total people} = 1000$$

$$P(\text{Actually Positive} \mid \text{positive test})$$

$$= \frac{P(\text{positive test} \mid \text{Actually Positive}) \times P(\text{Actually Positive})}{P(\text{positive test})}$$

$$= \frac{0.85 \times 0.70}{P(\text{positive test})} = 0.85 = \frac{0.70n}{P(\text{positive test})}$$

$$P(\text{positive test}) = \frac{0.70x}{0.85} \quad \textcircled{B}$$

$$P(\text{Actually positive} \mid \text{negative test}) = 0.10$$

$$0.10 = \frac{P(\text{negative test} \mid \text{Actually positive}) \times P(\text{Actually positive})}{P(\text{negative test})}$$

$$0.10 = \frac{(1-x) \times 0.70}{1 - P(\text{positive test})}$$

$$0.10 = \frac{(1-x) \times 0.70}{1 - \frac{0.70x}{0.85}} \quad \text{or, } 0.10 \left(1 - \frac{0.70x}{0.85} \right) = (1-x) \times 0.70$$

$$\text{or, } 0.085 - 0.70x = 0.595 - 0.595x$$

$$\text{or, } x = \frac{0.54}{0.525} \approx 0.9714$$

$$\therefore P(\text{positive test} \mid \text{Actually positive}) = 0.9714$$

$$P(\text{positive test}) = \frac{0.70 \times (0.9714)}{0.85}$$

$$= 0.8$$

$$\therefore P(\text{positive test}) = 0.80$$

$$\text{total population} = 1000$$

$$\therefore \text{people tested positive} = (1000 \times 0.80)$$

$$= 800 \text{ people.}$$

Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present, but has a 1% false positive rate (including that the virus is present, when it's not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two test use independent methods of identifying the virus. The virus is carried by 1% of all the people. Say that a person is tested for the virus using only one of the tests. And the test comes back positive for carrying the virus.

Which test returning positive is more indicative of someone really carrying the virus?

$$P(\text{virus} | \text{positive}) = \frac{P(\text{positive} | \text{virus}) P(\text{virus})}{P(\text{positive})}$$

$$P(\text{positive}) = P(\text{positive} | \text{virus}) P(\text{virus}) + P(\text{positive} | \text{No virus}) P(\text{No virus})$$

For test A:

$$\frac{0.95 \times 0.01}{(0.95 \times 0.01) + (0.1 \times 0.99)} = 0.0876$$

For test B:

$$\frac{0.90 \times 0.01}{(0.90 \times 0.01) + (0.05 \times 0.99)} = 0.1538$$

Linear Regression

$$Y = a_0 + a_1 x \quad \text{[same]}$$

$$Y = b + mx$$

Linear Regression equation is

given by $Y = a_0 + a_1 x + \text{error}$

where,

$$a_1 = \frac{(\bar{X}Y) - (\bar{X})(\bar{Y})}{\bar{X}^2 - (\bar{X})^2}$$

$$a_0 = \bar{Y} - a_1 \bar{X}$$

x_i (weeks)	\bar{Y}_i (Sales in thousands)
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

x_i (weeks)	(Sales in thousands)	x_i^2	$x_i \bar{Y}_i$	Prediction
1	1.2	1	1.2	1.2
2	1.8	4	3.6	1.86
3	2.6	9	7.8	2.52
4	3.2	16	12.8	3.18
5	3.8	25	19	3.84
sum	12.6	55	44.4	
Avg	2.52	11	8.88	

$$\text{Now, } \bar{x} = 3, \bar{Y} = 2.52, \bar{x}^2 = 11, \bar{xy} = 8.88$$

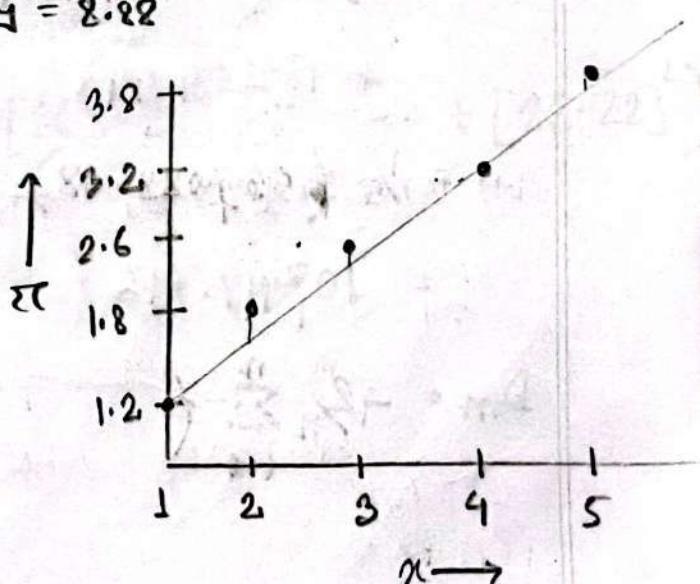
$$a_1 = \frac{(8.88) - (3 \times 2.52)}{(11) - (3)^2} = 0.66$$

$$a_0 = 2.52 - (0.66 \times 3) = 0.54$$

$$\begin{aligned} \therefore Y &= a_0 + a_1 x \\ &= 0.54 + 0.66x \end{aligned}$$

$$\text{For, } x = 7$$

$$\begin{aligned} Y &= 0.54 + 0.66 \times 7 \\ &= 5.16 \end{aligned}$$



Linear Regression Gradient Descent

$$\text{Loss function} = \sum (y_{\text{actual}} - \hat{y}_{\text{pred}})^2$$

$$\text{Mean Square Error (MSE)} = \frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i)^2$$

x _i	y _i
1	1.2
2	1.8
3	2.6
4	3.2
5	3.8

lets assume that,

in $y = mx + b$,

$m = 10, b = 300, \text{Learning Rate} = 0.001$

$$y = 10x + 300$$

$$\text{MSE} = \frac{1}{n} \sum_{i=0}^n (y_i - \hat{y}_i)^2 \quad [n = \text{total row}]$$

$$= \frac{1}{n} \sum_{i=0}^n (y_i - [mx + b])^2$$

$$\text{MSE} = \frac{1}{5} \sum (1.2 - 310)^2 + (1.8 - 320)^2 + (2.6 - 330)^2 + (3.2 - 340)^2 + (3.8 - 350)^2$$

$$= \frac{1}{5} \sum (95557.4 + 101251.24 + 107190.76 + 113434.24 + 119854.44)$$

$$= \frac{1}{5} (537088.08)$$

$$= 107417.616$$

$$Dm = -\frac{2}{n} \sum_{i=0}^n ($$

Problem Solving

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

Assume that, $\theta_0 = 1$ and $\theta_1 = 1.5$

Cost Function:

$$\frac{1}{2m} \sum_{i=1}^m (h_\theta(x_i) - y_i)^2$$

Gradient Decent

$$\frac{1}{2m} \sum_{i=1}^m (h_\theta(x_i) - y_i)^2$$

For, θ_0

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x_i) - y_i)$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x_i) - y_i) \times x_i$$

Cost Function

$$\frac{1}{2 \times 14} \sum ([11.5 - 11]^2 + [4 - 4]^2 + [26.5 - 25]^2 + \dots + [22 - 22]^2)$$

$$\frac{1}{28} \times 7.2025$$

$$= 0.25366071$$

X	Y	Pred-Y
7	11	11.5
21	4	4
17	25	26.5
9	15	14.5
4	7	7
11	16.5	17.5
12	19	19
6	10.2	10
1	2.3	2.5
3	5.1	5.5
2.5	4	4.75
3.8	7	6.7
9.6	11	9.4
14	22	22

1st Iteration

Assume that $\theta_0 = 1$ and $\theta_1 = 1$

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_0(x_i) - y_i)^2$$

$$\theta_0 = \theta_0 - \alpha \times \left[\frac{1}{14} (8-11) + (3-4) + (18-25) + \dots + (15-22) \right]$$

$$\theta_0 = \theta_0 - \alpha \times \left[\frac{1}{14} \times -47.2 \right]$$

$$\theta_0 = 1 - 0.01 \times (-3.3714) = 1.0337$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_0(x_i) - y_i) x_i$$

$$= 1 - \frac{0.01}{14} [8(8-11) + (3-4) \times 2 + (18-25) \times 17 + \dots]$$

$$= 1 - 0.01 \times (-33.6821) = 1.336$$

x_i	y_i	Pred- y
8	11	8
3	4	3
17	25	18
9	15	10
4	7	5
11	16.5	12
12	19	13
6	10.2	7
1	2.3	2
3	5.1	4
2.5	4	3.5
3.8	7	4.8
5.6	11	6.6
14	22	15

$$2508.7 \times \frac{1}{88}$$

$$18022.82 \times 0 =$$

Neural Network

Probability Problem

1. In the table below, you are given a dataset containing 9 rows and 3 features. Using naive bayes determine the most likely value of Y if $X_1 = 1, X_2 = a, X_3 = q$.

X_1	X_2	X_3	$P(Y=0 X_1=1, X_2=a, X_3=q)$
1	a	p	0
2	b	r	1
3	b	p	1
3	c	q	1
2	c	r	0
1	b	r	1
2	a	p	0
3	a	r	1
3	b	q	0

$$P(Y=0 | X_1=1, X_2=a, X_3=q)$$

For $Y=0$, rows = 0, 4, 6, 8

For $Y=1$, rows = 1, 2, 3, 5, 7

$$P(X_1=1 | Y=0) = \frac{1}{4}$$

$$P(X_2=a | Y=0) = \frac{2}{4} = \frac{1}{2}$$

$$P(X_3=q | Y=0) = \frac{1}{4}$$

$$P(Y=0) = \frac{4}{9}$$

$$P(Y=1) = \frac{5}{9}$$

$$1 = (0)9 + (1)9 + (1)9$$

$$1 = (0)9 + 8 \cdot 0 + 2 \cdot 0$$

$$1 = 0 + 0 + 0$$

$$1 = (0)9 + (1)9 = (1)9$$

$$1 = (0)9 + (1)9 = (1)9$$

$$P(Y=0 \mid X_1=1, X_2=a, X_3=q) = \frac{4}{9} \times \frac{1}{5} \times \frac{1}{4} \times \frac{2}{4} \times \frac{1}{4}$$

$$= \frac{1}{72} \approx 0.0138$$

Again, ~~from 2.6.2.2, joint probability~~ ~~from 2.6.2.2, joint probability~~ ~~from 2.6.2.2, joint probability~~

$$P(X_1=1 \mid Y=1) = \frac{1}{5} \quad P(X_2=a \mid Y=1) = \frac{1}{5}$$

$$P(X_3=q \mid Y=1) = \frac{2}{5}$$

$$\therefore P(Y=1 \mid X_1=1, X_2=a, X_3=q) = \frac{5}{9} \times \frac{1}{5} \times \frac{2}{5} \times \frac{1}{5}$$

$$= \frac{2}{225} \approx 8.889 \times 10^{-3}$$

$$\therefore 0.0138 > 8.889 \times 10^{-3}$$

$$\therefore Y=0 \quad [Ans]$$

$$2. X = \{A, B, C\}. \text{ Assume } P(A) = 0.5, P(B) = 0.3.$$

Determine $P(C)$ and $P(A \cup B)$

We know that,

$$P(A) + P(B) + P(C) = 1$$

$$\text{or, } 0.5 + 0.3 + P(C) = 1$$

$$\text{or, } P(C) = 0.2.$$

$$P(A \cup B) = P(A) + P(B) = 0.5 + 0.3 = 0.8$$

$$= (0=Y \mid A \cup B = \{X\})$$

$$= (0=Y \mid P=ex)$$

3. Assume 2 coins are tossed simultaneously.

Event A = The 1st coin coming up head.

Event B = The 2nd coin coming up tail.

What is $P(A \cap B)$?

Sample Space, S = {HH, HT, TH, TT}

When the first coin is head, {HH, HT}

When the 2nd coin is tail, {HT, TT}

$\therefore P(A \cap B) = \{HT\}$ | Total space = 4, output = 1

$\therefore P(A \cap B) = \frac{1}{4} = 0.25$.

4. (i) Assume Event A & B are

absolutely independent and $P(B|A) = 0.5$

What is the value of x?

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B \cap A)}{P(A)} = 0.5$$

Now,

$$\frac{P(A \cap B)}{P(A)} = 0.5 \quad \text{from the table, } P(B \cap A) = 0.1$$

$$\therefore P(A) = \frac{0.1}{0.5} = 0.2$$

~~$$P(A) = P(B) + P(B')$$~~

~~$$0.2 = 0.1 + x + 0.2 + 0.1$$~~

~~$$\therefore x = -0.2$$~~

	A		A'	
	B	B'	B	B'
C	0.1	0.2	0.2	0.1
C'	x	0.1	0.1	0.1

Given that,

$$P(B|A) = 0.5$$

$$P(A \cap B) = P(C \cap A \cap B) + P(C' \cap A \cap B)$$

$$= 0.1 + X$$

$$P(A) = P(C \cap A \cap B) + P(C \cap A \cap B') + P(C' \cap A \cap B) + P(C' \cap A \cap B')$$

$$= 0.1 + 0.2 + X + 0.1$$

$$= 0.4 + X$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.1 + X}{0.4 + X} = 0.5$$

$$0.5 = \frac{0.1 + X}{0.4 + X}$$

$$0.5 = \frac{1}{P} = (A \cap B)^{-1}$$

$$\text{or, } 0.2 + 0.5X = 0.1 + X$$

$$\text{or, } 0.2 - 0.1 = X - 0.5X$$

$$\text{or, } 0.1 = 0.5X$$

$$\text{or, } X = 0.2$$

b) using $X = 0.2$ determine Y .

$$0.1 + 0.2 + 0.2 + Y + 0.2 + 0.1 + 0.1 + 0.1 = 1$$

$$\text{or, } 1 + Y = \frac{1.0}{2.0} = (A)^{-1}$$

$$\text{or, } Y = 0$$

$$\therefore Y = 0$$

$$(A)^{-1} + (A)^{-1} = (A)^{-1}$$

$$1.0 + 0.0 + X + 1.0 = 2.0$$

① Determine $P(A|B \cap C)$

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

$$= \frac{0.1}{0.3} = 0.333$$

Here, $P(A \cap B \cap C) = 0.1$

$$P(B \cap C) = 0.1 + 0.2 = 0.3$$

	Pandemic	No Pandemic	Total		
	Online	Offline	Online	Offline	Total
Public University	0.142	0.037	0.165	0.072	
Private University	0.102	0.145	0.217	0.118	
Total	0.244	0.182	0.382	0.190	

② Is pandemic Conditionality Independent of Public University Given that Online Class?

For Conditional Independence:

$$P(A|B, C) = P(A|C)$$

given that part

$$P(\text{Pand} | \text{Pub, Online}) = \frac{P(\text{Pand} \cap \text{Pub} \cap \text{online})}{P(\text{Pub} \cap \text{online})}$$

$$= \frac{0.142}{0.142 + 0.165}$$

$$= 0.46254$$

$$= 0.391$$

Not conditionally independent.

$$P(\text{Pub} | \text{Pand} | \text{online}) = \frac{P(\text{Pand} \cap \text{online})}{P(\text{online})} = \frac{0.142 + 0.103}{0.627}$$

④ Is Private Or Independent of Online Class?

$$P(\text{Private} \cap \text{Online}) = P(\text{Private}) \cdot P(\text{Online})$$

$$P(\text{Private} \cap \text{Online}) = 0.103 + 0.07 = 0.82 \quad \text{--- (a)}$$

$$P(\text{Private}) = 0.103 + 0.146 + 0.217 + 0.118 = 0.584$$

$$P(\text{Online}) = 0.142 + 0.103 + 0.165 + 0.217 = 0.627$$

$$P(\text{Private}) \cdot P(\text{Online}) = 0.584 \times 0.627$$

$$= 0.366168 \quad \text{--- (b)}$$

Not Independent

④ Find the marginal probability of offline class

$$P(\text{Public}) \rightarrow (0.037 + 0.146 + 0.072 + 0.118)$$

$$= 0.373$$

$$\frac{P(\text{Public} \cap \text{Online})}{P(\text{Public})} = \frac{P(\text{Online} | \text{Public})}{P(\text{Public})}$$

$$\frac{0.072}{0.373}$$

$$= \frac{0.191}{0.373}$$

$$= 0.510$$

$$P(\text{Public} \cap \text{Online}) =$$

Marginal probability
of Online class

$$0.072 + 0.118$$

$$= (0.190 | \text{Public})$$

$$= (0.190 | \text{Online})$$

$$= 0.510$$

$$= (0.190 | \text{Online})$$

$$\textcircled{a} \quad h_w(x) = w_1 x + w_0$$

here compare the equation with $y = mx + c$

$$m = w_1 \text{ and } c = w_0$$

w_1 = slope of the line.

w_0 = the intercept point.

\textcircled{b} given that,

$$\text{Loss}(h_w) = \sum_{i=1}^N (y_i - (w_1 x_i + w_0))^2$$

$$\nabla \rightarrow \frac{\partial \text{Loss}}{\partial w_1} = -2 \sum x_i (y_i - h_w(x_i))$$

$$\nabla \rightarrow \frac{\partial \text{Loss}}{\partial w_0} = -2 \sum (y_i - h_w(x_i))$$

Initially $w_1 = 15, w_0 = 0$

Iteration	$h(w)$	Loss	x	y	w_1	w_0
1	$-410.8x - 132$	-66	132	425.8	-132	-410.8
2	$31495.3x - 11726.2$	5959.08	-11918.16	-31906.096	11786.16	31495.3

$$h_w(x) = 15x + 0$$

Here, error = Actual - Prediction

$$h_{w_1}(x) = 15(1.6) = 24$$

$$e_1 = (30 - 24) = 6$$

$$h_{w_2}(x) = 15(2.0) = 30$$

$$e_2 = (27 - 30) = -3$$

$$h_{w_3}(x) = 15(2.5) = 37.5$$

$$e_3 = (24 - 37.5) = -13.5$$

$$h_{w_4}(x) = 15(3.0) = 45$$

$$e_4 = (22 - 45) = -23$$

$$h_{w_5}(x) = 15(3.5) = 52.5$$

$$e_5 = (20 - 52.5) = -32.5$$

$$\frac{\partial \text{Loss}}{\partial w_0} = -2 \sum (y_i - h_w(x))$$

$$= -2(6 - 3 - 13.5 - 23 - 32.5) \\ = -132$$

$$\frac{\partial \text{Loss}}{\partial w_1} = -2 \sum x_i (y_i - h_w(x))$$

$$= -2[(1.6 \times 6) + (2.0 \times -3) + (2.5 \times -13.5) + (3 \times -23) + (3.5 \times -32.5)] \\ = -2[9.6 - 6.0 - 33.75 - 69 - 113.75] \\ = -2(-212.9) \\ = 425.8$$

Now,

$$w_0 = w_0 - \alpha \frac{\partial \text{Loss}}{\partial w_0}$$

$$= 0.1 \times (-132)$$

$$= -13.2$$

$$w_1 = w_1 - \alpha \frac{\partial \text{Loss}}{\partial w_1}$$

$$= 0.1 \times (425.8)$$

$$= 42.58$$

Final equation: $-410.8x - 13.2$

Again,

$$h_w(x) = -410.8x - 132$$

$$h_{w1}(x) = -410.8(1.6) - 132 = -729.28$$

$$h_{w2}(x) = -410.8(2.0) - 132 = -953.6$$

$$h_{w3}(x) = -410.8(2.5) - 132 = -1159$$

$$h_{w4}(x) = -410.8(3.0) - 132 = -1364.4$$

$$h_{w5}(x) = -410.8(3.5) - 132 = -1569.8$$

Error = Actual - Prediction Value

$$e_1 = (20 + 729.28) = 819.28$$

$$e_2 = (27 + 953.6) = 980.6$$

$$e_3 = (24 + 1159) = 1183$$

$$e_4 = (22 + 1364.4) = 1386.4$$

$$e_5 = (20 + 1569.8) = 1589.8$$

$$\frac{\partial L_{\text{loss}}}{\partial w_0} = -2 \sum (y_i - h_w(x))$$

$$= -2(819.28 + 980.6 + 1183 + 1386.4 + 1589.8)$$

$$= -11918.16$$

$$\frac{\partial L_{\text{loss}}}{\partial w_1} = -2 \sum x (y_i - h_w(x))$$

$$= -2 [(1.6 \times 819.28) + (2.0 \times 980.6) + (2.5 \times 1183) + (3.0 \times 1386.4) + (3.5 \times 1589.8)]$$

$$= -2 [1310.898 + 1961.2 + 2957.5 + 4159.2 + 5564.3]$$

$$= -31906.096$$

Now,

$$w_0 = w_0 - \alpha \frac{\partial \text{Loss}}{\partial w_0}$$

$$= -132 - 1(-11786.16)$$

$$= 31786.16$$

$$w_1 = w_1 - \alpha \frac{\partial \text{Loss}}{\partial w_1}$$

$$= -410.8 - 1(-31906.096)$$

$$= -410.8 + 31906.096$$

$$= 31495.296$$

Final equation: $31495.296x + 11786.16$

③ from question 2;

$$31495.296x + 11786.16$$

$$\text{for } x=3, y = 31495.296(3) + 11786.16$$

$$= 106272.048 \text{ mpg}$$

Actual = 22 mpg

Predict = 106272.048 mpg

$$\text{difference} = (106272.048 - 22) \text{ mpg}$$

$$= 106250.048 \text{ mpg}$$

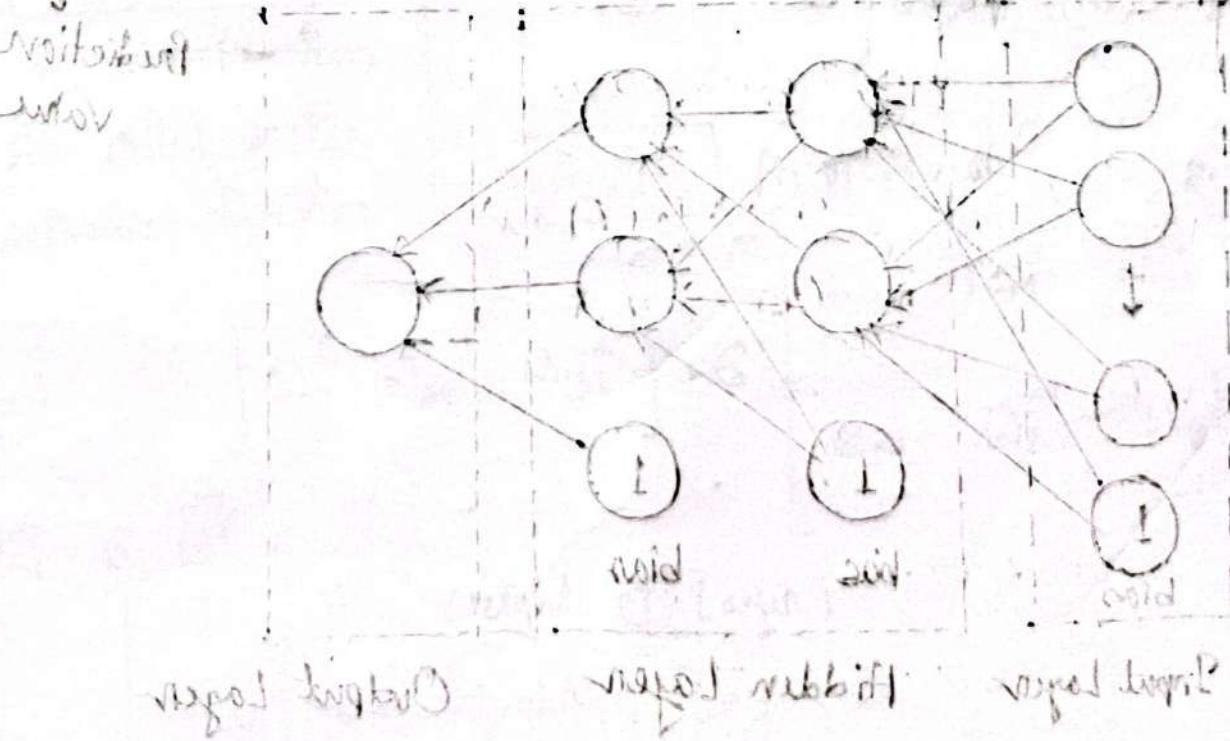
Here learning rate $\alpha=1$ causes a misslead of prediction. The rate must in range of 0.01 to 0.001 for better output.

④

Stochastic Gradient

Large learning rate causes of divergence of the output. Also a large learning rate may cause a bad prediction. On the other hand, a decaying learning rate start with a high value to jump quickly and slowly reduce over-time.

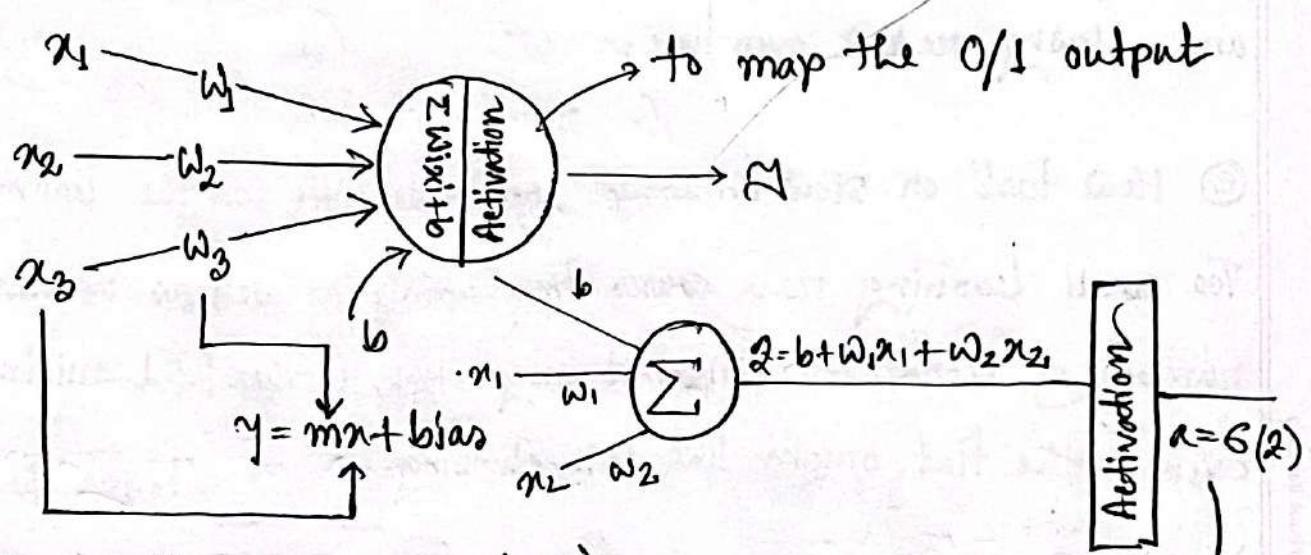
⑤ How fast or slow a model learns depends on the learning rate. Too small learning rate causes tiny update as well as increase the number of iteration. Also, it may stuck in the local minima or even in the flat origin like hill climbing.



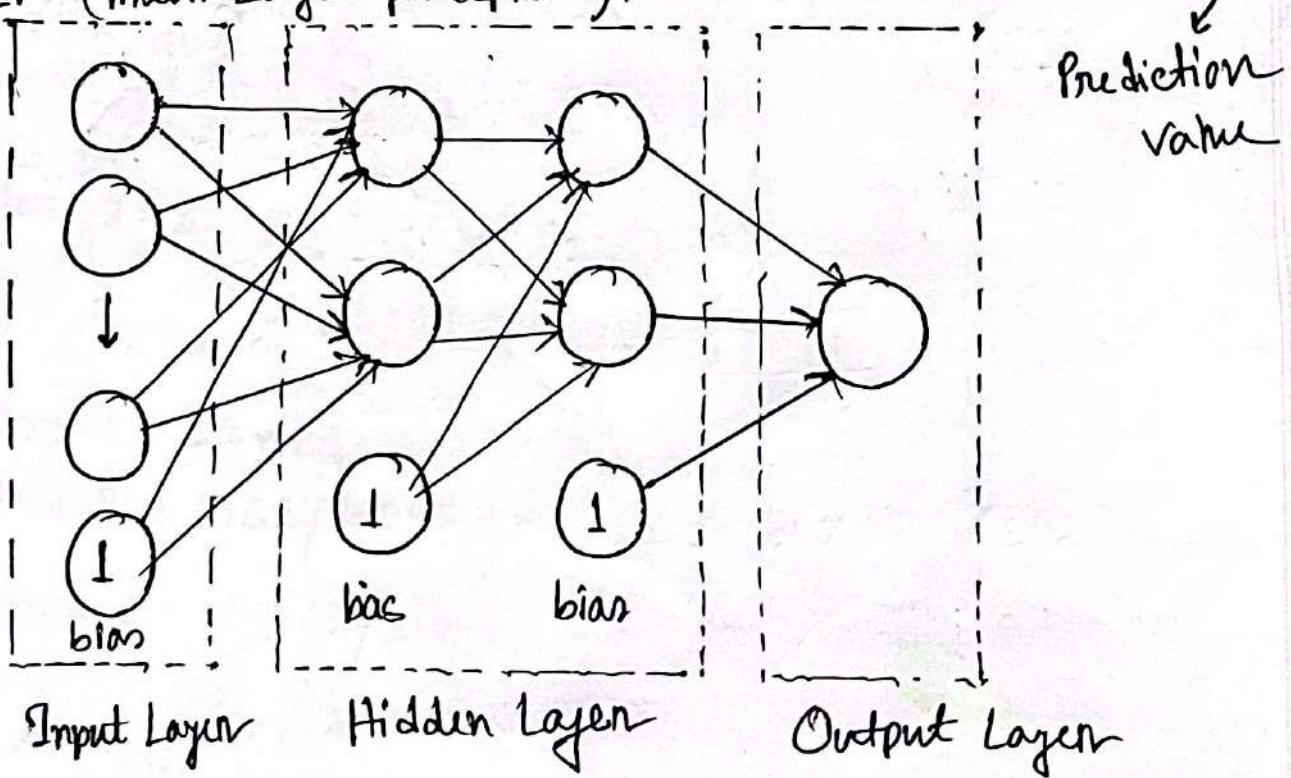
Neural Network

What is a perceptron?

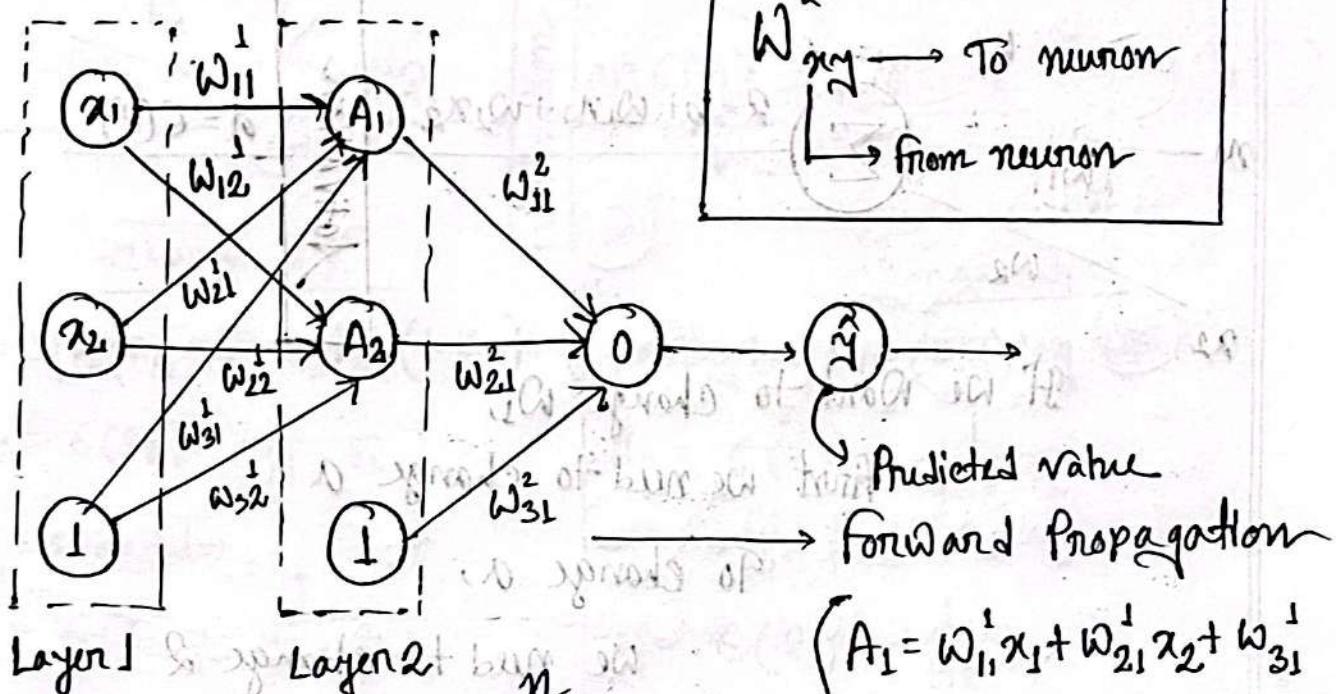
One layer neural networks that can classify things into two parts.



MLP (Multi Layer perceptron):



Forward Propagation



Layer 1 Layer 2

$$f(x) = \frac{1}{1 + e^{-x}}$$

Sigmoid function

Also called as the activation function

$$w_{in} + b = \sum_{i=1}^n w_{in}^i x_i + b$$

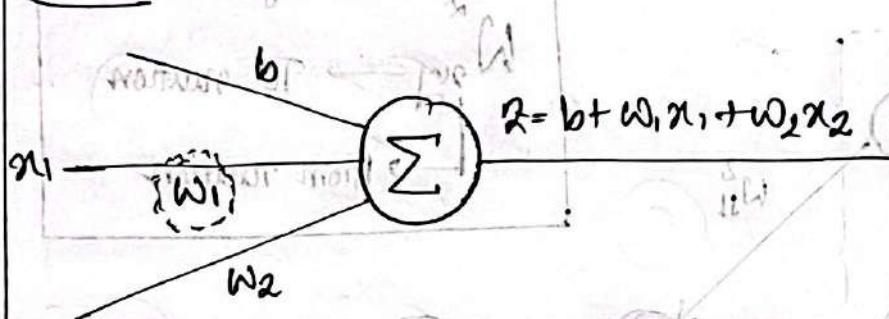
$$\begin{cases} A_1 = w_{11}^1 x_1 + w_{21}^1 x_2 + w_{31}^1 \\ A_2 = w_{12}^1 x_1 + w_{22}^1 x_2 + w_{32}^1 \\ 0 = w_{11}^2 f(A_1) + w_{21}^2 f(A_2) + w_{31}^2 \end{cases}$$

$$\hat{z} = G \cdot \begin{bmatrix} w_{11}^2 & w_{21}^2 & w_{31}^2 \end{bmatrix} \cdot G$$

$$\begin{bmatrix} w_{11}^1 & w_{21}^1 & w_{31}^1 \\ w_{12}^1 & w_{22}^1 & w_{32}^1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

$$\hat{z} = G \cdot w^2 \cdot G^T \cdot w^1 \cdot x$$

Chain Rules:



activation function

If we want to change w_1 ,
first we need to change a
To change a ,

we need to change z first

According to the chain rules:

$$\frac{dL}{dw_1} = \frac{dL}{da} \times \frac{da}{dz} \times \frac{dz}{dw_1}$$

$$L = -y \log a - (1-y) \log (1-a)$$

LOSS

Why this dependency?

$$L = -y \log a - (1-y) \log (1-a)$$

$$a = g(z) = \frac{1}{1+e^{-z}}$$

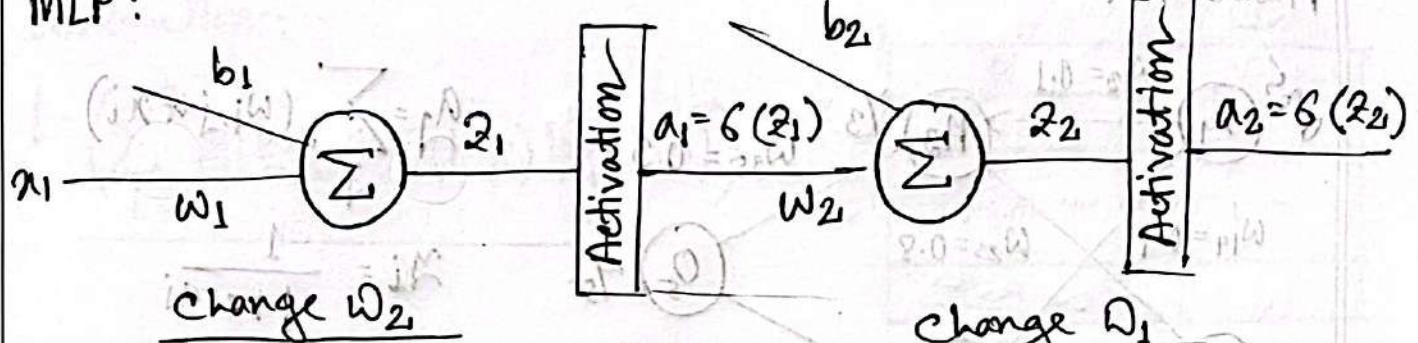
$$z = b + x_1 w_1 + x_2 w_2$$

input to hidden

input to hidden

activation function

MLP:



$$L = -y \log a_2 - (1-y) \log (1-a_2)$$

$$a_2 = G(z_2) = \frac{1}{1+e^{-z_2}}$$

$$z_2 = a_1 w_2 + b_2$$

$$a_1 = G(z_1) = \frac{1}{1+e^{-z_1}}$$

$$z_1 = x_1 w_1 + b_1$$

$$\frac{dL}{dw_2} = \frac{d z_2}{d w_2} \times \frac{d a_2}{d z_2} \times \frac{dL}{d a_2}$$

w_1

$$\begin{aligned} \frac{dL}{dw_1} &= \frac{d z_1}{d w_1} \times \frac{d a_1}{d z_1} \times \frac{d z_2}{d a_1} \times \frac{d a_2}{d z_2} \times \frac{dL}{d a_2} \\ &= (0.5 \times 2.0) + (1.0 \times 0.5) = 1.0 \\ &= (0.5 \times 0.0) + (1.0 \times 0.0) = 0.0 \end{aligned}$$

w_2

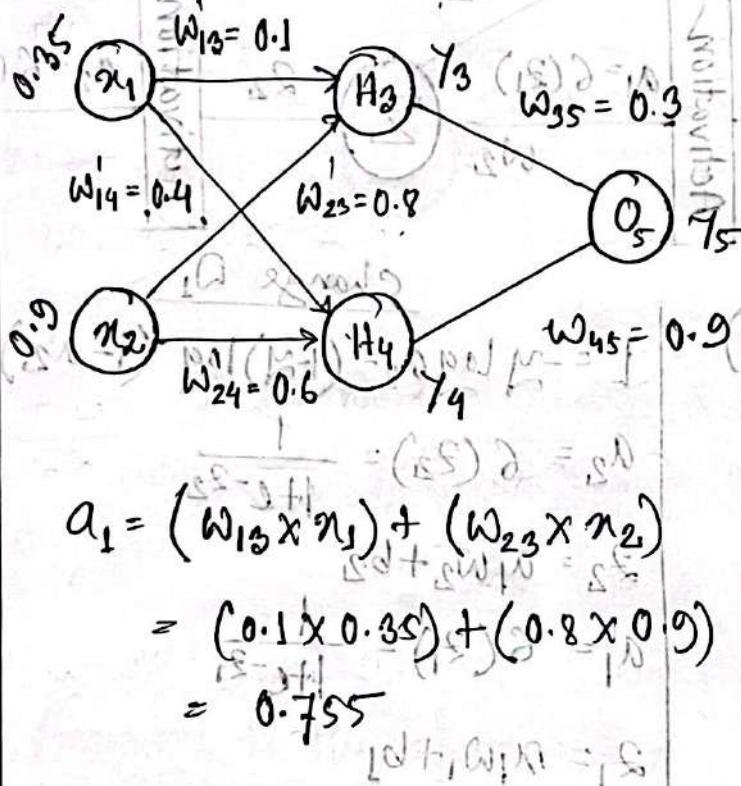
Here w_1 and w_2 are independent. Change in w_1 does not effect in w_2 . But does effect a_1 and z_1 .

For w_1 , w_2 & w_3 :

$$\frac{dL}{dw_1} = \frac{d z_1}{d w_1} \times \frac{d a_1}{d z_1} \times \frac{d z_2}{d a_1} \times \frac{d a_2}{d z_2} \times \frac{d z_3}{d a_2} \times \frac{d a_3}{d z_3} \times \frac{dL}{d a_3}$$

Problems:

: 9/11



$a_2 = (w_{14} \times x_1) + (w_{24} \times x_2)$
 $= (0.4 \times 0.35) + (0.6 \times 0.9)$
 $= 0.68$

$a_3 = (w_{35} \times y_3) + (w_{45} \times y_4)$
 $= (0.3 \times 0.68) + (0.9 \times 0.6637)$

$y_5 = 0.801$

$y_5 = \frac{1}{1+e^{-x}} = \frac{1}{1+e^{-0.801}} = 0.69$ [Output]

Error: $y_{\text{target}} - y_{\text{predicted}} = (0.5 - 0.69) = -0.19$

$$a_i = \sum_{j=1}^n (w_{i,j} \times x_j)$$

$$y_i = \frac{1}{1+e^{-a_i}}$$

$$y_5 = 0.5$$

$$a_3 = \frac{1}{1+e^{-0.755}} = 0.68$$

$$y_4 = 0.68$$

$$a_2 = 0.68$$

$$y_2 = \frac{1}{1+e^{-0.68}} = 0.6637$$

IC

Loss Function:

$$L = \frac{\left(\sum_{i=1}^3 -y_i \log a_i - (1-y_i) \log (1-a_i) \right)}{3}$$

X	Y
0.1	0
0.2	0
0.3	1
0.4	1

Assume that,

$$w = 0.7 \text{ and } b = 0.1$$

$$Z = 0.7x + 0.1 \quad [Y = mx + b]$$

$$\begin{aligned}
 L &= -0 \log \left(\frac{1}{1+e^{-(0.7 \times 0.1 + 0.1)}} \right) - (1-0) \log \left(1 - \frac{1}{1+e^{-(0.7 \times 0.1 + 0.1)}} \right) \\
 &\quad - \boxed{} + \boxed{} \\
 &= -\log \left(1 - \frac{1}{1+e^{-(0.7 \times 0.1 + 0.1)}} \right) - \log \left(1 - \frac{1}{1+e^{-(0.7 \times 0.2 + 0.1)}} \right) \\
 &\quad - \log \left(1 - \frac{1}{1+e^{-(0.7 \times 0.3 + 0.1)}} \right) + \\
 &= -\log (1-0.54) - \log (1-0.55) - \log 0.57 \\
 &= 0.928
 \end{aligned}$$

$$\text{Avg corr value} = \left(\frac{0.928}{3} \right) = 0.309$$

$$\sqrt{\left(1 - \frac{1}{1+e^{-(0.7 \times 0.1 + 0.1)}} \right) + \left(1 - \frac{1}{1+e^{-(0.7 \times 0.2 + 0.1)}} \right) + \left(1 - \frac{1}{1+e^{-(0.7 \times 0.3 + 0.1)}} \right)}$$

$$\frac{dL}{dw_i} = \left(\sum_{i=1}^n \left(\frac{1}{1+e^{-x_i}} - y_i \right) x_i \right) \quad \text{for } i=1, 2, 3$$

$$w = w - \alpha \frac{dL}{dw}$$

$$\frac{dL}{db} = \left(\sum_{i=1}^n \left(\frac{1}{1+e^{-x_i}} - y_i \right) \right) \quad \text{for } i=1, 2, 3$$

$$b = b - \alpha \frac{dL}{db}$$

$$\frac{dL}{dw} = \left(\left(\frac{1}{1+e^{-(0.7 \times 0.1 + 0.1)}} - 0 \right) 0.1 + \left(\frac{1}{1+e^{-(0.7 \times 0.2 + 0.1)}} - 0 \right) 0.2 \right. \\ \left. + \left(\frac{1}{1+e^{-(0.7 \times 0.3 + 0.1)}} - 1 \right) 0.3 \right) / 3 = 0.013$$

$$= \frac{(0.054 + 0.112 - 0.127) / 3}{3} = 0.013$$

$$= 0.013$$

$$\frac{dL}{db} = \left(\left(\frac{1}{1+e^{-(0.7 \times 0.1 + 0.1)}} - 0 \right) + \left(\frac{1}{1+e^{-(0.7 \times 0.2 + 0.1)}} - 0 \right) \right. \\ \left. + \left(\frac{1}{1+e^{-(0.7 \times 0.3 + 0.1)}} - 0 \right) \right) / 3 = 0.013$$

$$= \frac{(0.54 + 0.56 - 0.42)}{3}$$

$$= 0.23$$

if learning rate, $\alpha = 1$ i.e. $(\text{loss}_1 - \text{loss}_0)$

$$\text{updated weight, } w = w - \alpha \frac{dL}{dw}$$

$$= 0.7 - 1(0.013)$$

$$w = 0.687 \quad \Rightarrow \quad \frac{10}{w_0}$$

$$\text{updated bias, } b = b - \alpha \frac{dL}{db}$$

$$= 0.1 - 1(0.23) \quad \Rightarrow \quad \frac{10}{b_0}$$

$$= -0.13$$

$$\text{old } Q = 0.7w + 0.1$$

$$\text{new } Q = 0.687w - 0.13$$

$$L \text{ for new } Q \text{ will be, } L = 0.86 \quad \Rightarrow \quad \frac{10}{L_0} \quad \text{old value}$$

$$\text{Avg loss value} = \frac{0.86}{3}$$

$$= 0.2866 \quad [\text{new}]$$

$$\text{Avg loss value} = 0.309 \quad [\text{previous}]$$

Conclusion: error/loss slightly reduce from 0.309 to 0.2866.

LOSS function:

$$(y_{pred} - y_{act})^2$$

Logistic Regression:

$$\frac{\partial \text{Loss}}{\partial w_j} = (y_{pred} - y_{act}) \cdot x_j$$

$$\frac{\partial \text{Loss}}{\partial b} = (y_{pred} - y_{act})$$

$$\frac{\partial L}{\partial w} = \sum_{i=1}^n \left(\frac{1}{1+e^{-x_i \cdot w}} - y_i \right) x_i$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \left(\frac{1}{1+e^{-x_i \cdot b}} - y_i \right)$$

Update weight:

$$w = w - \alpha \frac{\partial \text{Loss}}{\partial w}$$

Update bias:

$$w = w - \alpha \frac{\partial \text{Loss}}{\partial b}$$

Sigmoid function:

$$\frac{d}{dx} \left(\frac{1}{1+e^{-x}} \right) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$\text{or, } G'(x) = G(x) \cdot (1 - G(x))$$

Gradient Descent Loss Function:

$$L = \frac{1}{n} \sum_{i=1}^n (y_{\text{act}} - (w_0 + w_1 x_i))^2$$

$$\frac{\partial L}{\partial w_0} = -\frac{2}{n} \sum_{i=1}^n x_i (y_{\text{act}} - y_{\text{pred}})$$

$$\frac{\partial L}{\partial w_1} = -\frac{2}{n} \sum_{i=1}^n (y_{\text{act}} - y_{\text{pred}})$$

Actual Data Table:

>Create an equation and predict the output

Calculate the loss function

$$\text{Update } w \text{ and bias}$$

MSE/SSR:

Actual Table

$$w = w - \alpha \frac{\partial L}{\partial w}$$

$$b = b - \alpha \frac{\partial L}{\partial b}$$

Prediction

find Loss $\sum (actual - predict)^2$

$$\frac{1}{n}$$

$$\begin{aligned}
 E(\text{Decision}) &= -P(\text{Yes}) \log_2 P(\text{Yes}) - P(\text{No}) \log_2 P(\text{No}) \\
 &= -\frac{9}{14} \log_2 \left(\frac{9}{14}\right) - \frac{5}{14} \log_2 \left(\frac{5}{14}\right) \\
 &= 0.990
 \end{aligned}$$

(using P = 1 - No. in 3 M) $= \frac{16}{16} = \frac{16}{16}$

Students

$$E(\text{student}) = \frac{1}{10} \log_2 \left(\frac{1}{10}\right) + \frac{9}{10} \log_2 \left(\frac{9}{10}\right)$$

$$E(\text{Decision}) = -P(\text{Yes}) \log_2 P(\text{Yes}) - P(\text{No}) \log_2 P(\text{No})$$

$$\text{Ans} = \frac{4}{10} \log_2 \left(\frac{4}{10}\right) + \frac{6}{10} \log_2 \left(\frac{6}{10}\right)$$

$$= 0.971$$

using P = 1 - No. in 3 M

Ans

$$E(\text{Yes} | A_1) = \frac{4}{5} \log_2 \left(\frac{4}{5}\right) + \frac{1}{5} \log_2 \left(\frac{1}{5}\right) = \frac{3}{5} \log_2 \left(\frac{4}{5}\right)$$

$$\frac{4}{5} \log_2 \left(\frac{4}{5}\right) = 0.22$$

$$E(\text{No} | A_1) = -\frac{1}{5} \log_2 \left(\frac{1}{5}\right)$$

$$= 0.22 \text{ M}$$

using P = 1 - No. in 3 M

$$\begin{aligned}
 & \frac{4}{5} \log_2 \left(\frac{4}{5}\right) + \frac{1}{5} \log_2 \left(\frac{1}{5}\right) \\
 & \quad \text{using P = 1 - No. in 3 M}
 \end{aligned}$$

using P = 1 - No. in 3 M

$$\begin{aligned}
 E(D) &= -P(Y) \log_2 P(Y) - P(N) \log_2 P(N) \\
 &= -\frac{4}{10} \log_2 \left(\frac{4}{10}\right) - \frac{6}{10} \log_2 \left(\frac{6}{10}\right) \\
 &= 0.971
 \end{aligned}$$

 $x_1:$

$$\begin{aligned}
 E(x_1) &= -P(Y|x_1) \log_2 P(Y|x_1) - P(N|x_1) \log_2 P(N|x_1) \\
 &= -\frac{4}{5} \log_2 \left(\frac{4}{5}\right) - \frac{1}{5} \log_2 \left(\frac{1}{5}\right) = 0.72
 \end{aligned}$$

$$\begin{aligned}
 E(a_2) &= -P(Y|a_2) \log_2 P(Y|a_2) - P(N|a_2) \log_2 P(N|a_2) \\
 &= -\frac{0}{5} \log_2 \left(\frac{0}{5}\right) - \frac{5}{5} \log_2 \left(\frac{5}{5}\right) = 0.3610
 \end{aligned}$$

$$\begin{aligned}
 I(A) &= 0.971 - 0.72 P(a_1) - 0.361 P(a_2) \\
 &= 0.971 - 0.72 \frac{5}{10} - \cancel{0.361} \frac{5}{10} \\
 &= 0.610
 \end{aligned}$$