1

Name: Rollno: Signature:

CS 302 – Artificial Intelligence - EndSem 15th April, 2021 Be Concise, Weightage - 30%

Instructions*: 1. EndSem is closed-book. Cheatsheets, cellphones are not allowed. 2. Answers are to be written on the question paper. 3. Two page supplement may be asked from the invigilators for rough work.

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total

1. The algorithm A* is about to expand node N in Figure below. The directed arrows are parent pointers, and the numbers in the nodes are g-values. Edge labels where given are edge costs. Show the graph after A* has finished expanding node N. (**12 Marks**)

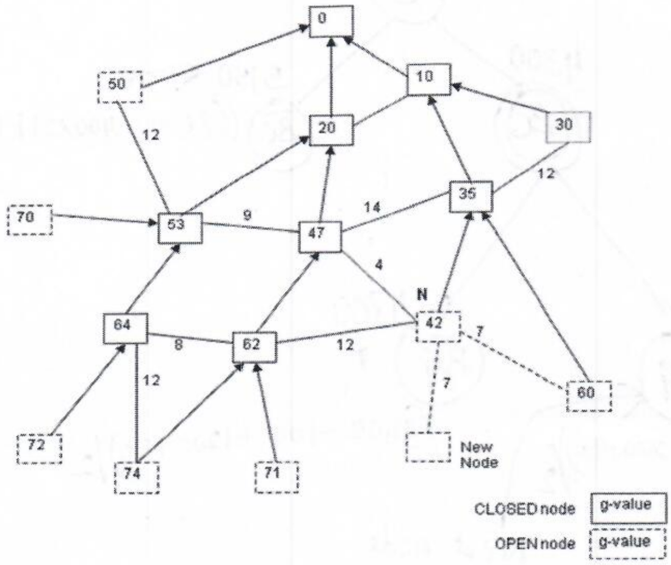
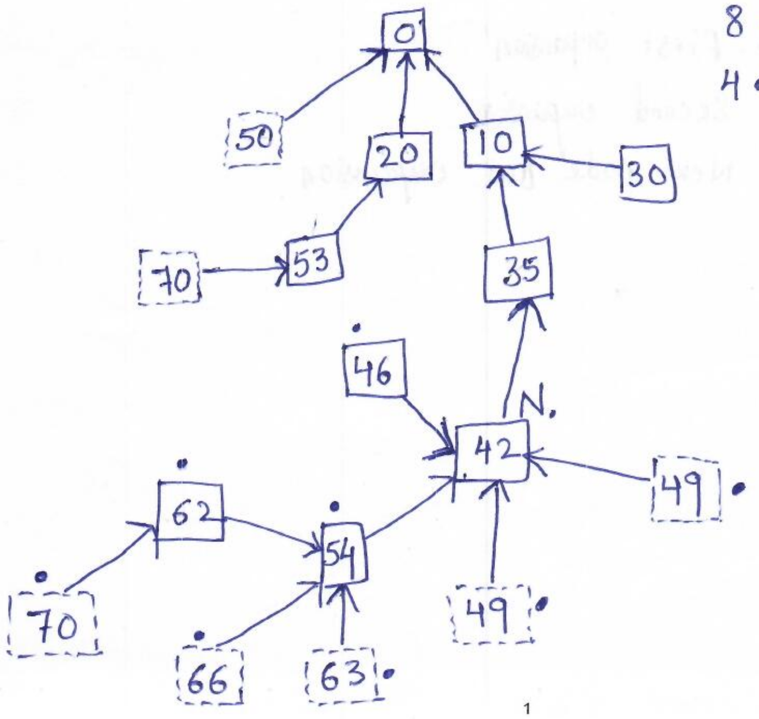


Figure 1: A* is about to expand Node N

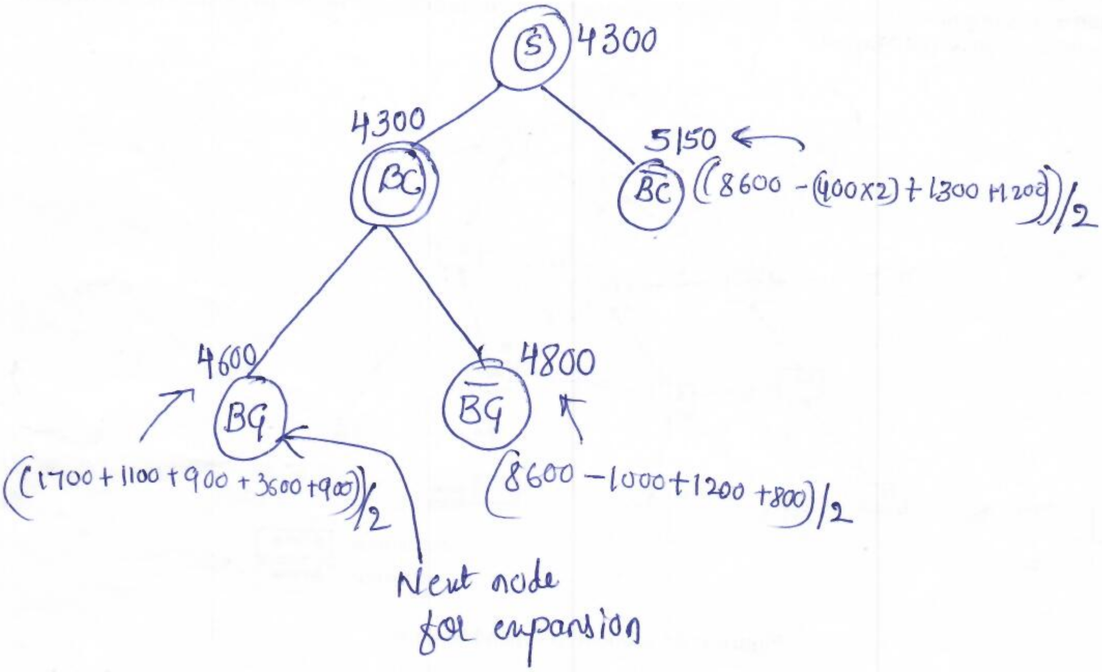
8 marks → 8 updated Node
4 marks → updated parent Pointers



2. Given the cost matrix below, show two expansions of the Branch and Bound algorithm in the solution space towards solving the TSP (traveling salesman problem). Use the heuristic "add cheapest allowed edge" when refining a node. Starting with the root label each node in your search tree with the lower bound estimated costs as discussed in the class. Mark the next node that is selected by the algorithm. (**10 marks**)

	Chennai	Goa	Mumbai	Delhi	Bangalore	
Chennai	0	800	1300	2200	400	1200
Goa	800	0	600	2100	500	1100
Mumbai	1300	600	0	1500	1200	1800
Delhi	2200	2100	1500	0	2400	3600
Bangalore	400	500	1200	2400	0	900
						8600

$\frac{8600}{2} = 4300$



4 marks → First expansion
5 marks → Second expansion
1 mark → Next node for expansion.

3. Draw arcs from left to right to relate each term on the LHS to all related terms on the RHS. (**8 Marks**)

1 → c, d, f (a, b)
2 → c, e, f
3 → h
4 → g
5 → h
6 → b, d, f
7 → e, g, h (a, b)
8 → g, h

a. Depth First	1. systematic
b. Breadth First	2. heuristic
c. Best First	3. crossover
d. Branch & Bound	4. temperature
e. Hill Climbing	5. population size
f. A*	6. admissible
g. Simulated Annealing	7. unsystematic
h. Genetic Algorithm	8. probabilistic

4. Given the Bayesian network shown in Figure 2, compute the following probabilities (show the steps): (**6 marks**)

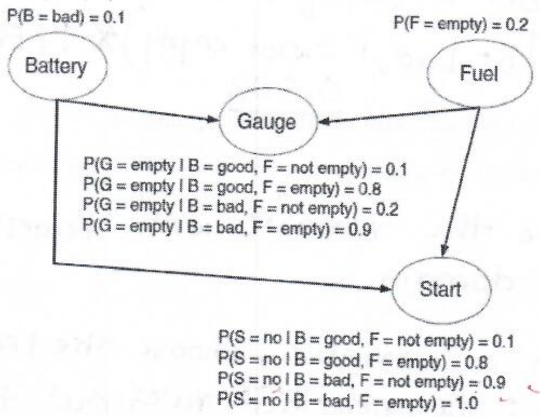


Figure 2: Bayesian belief network

(a) P(B = good, F = empty, G = empty, S = yes).

$P(B=good, F=empty, G=empty, S=yes)$

$= P(B=good) \times P(F=empty) \times P(G=empty \mid B=good, F=empty) \times P(S=yes \mid B=good, F=empty)$

$= 0.9 \times 0.2 \times 0.8 \times 0.2 = \boxed{0.0288}$

(b) P(B = bad, F = empty, G = not empty, S = no).

$= P(B=bad) \times P(F=empty) \times P(G=\text{not empty} \mid B=bad, F=empty) \times P(S=no \mid B=bad, F=empty)$

$= 0.1 \times 0.2 \times 0.1 \times 1.0 = \boxed{0.002}$

OR

$P(S=no \mid B=Bad, F=Empty) \times P(G=\text{not emp} \mid B=bad, F=empty) \times P(B=bad) \times P(F=empty)$

$= 1 \times (1-0.9) * 0.1 \times 0.2 = 0.1 \times 0.1 \times 0.2$

$= 0.002$

(c) Given that the battery is bad, compute the probability that the car will start.

$P(S=yes \mid B=bad)$ $\alpha \rightarrow$ empty or not

$= \sum_{\alpha} P(S=yes \mid B=bad, F=\alpha)\, P(B=bad)\, P(F=\alpha)$

$= 0.1 \times 0.1 \times 0.8 = \boxed{0.008}$

OR

$P(C=start \mid B=bad)$

$P(S=yes \mid B=bad, F=Empty) \times P(B=bad) \times P(F=Empty) +$

$P(S=yes \mid B=bad, F=not\ empty) * P(bad) + P(not\ empty)$

$= 0.008$

5. For Graph coloring problem (a type of constraint satisfaction problem)

(a) What is the use of the heuristic, Minimum Remaining Values and Least Constraining value? (**2 marks**)

MRV → Choose the variable with fewest legal values in its domain

LCV → Given a variable, choose the least constraining value : the one that rules out the fewest values in remaining variables

(b) How would you use forward checking (FC) in early detection of inevitable failures? What is its limitation? (**2 marks**)

FC → Keep track of remaining legal values for the unassigned variables. Terminate when any variable has no legal values.

Limitations →

① doesnot check interaction betn. unassigned variables.

② doesnot look very far in the future, hence doesnot detect all failures

(c) What is arc consistency and how it can be used for early detection of inevitable failures? (**2 marks**)

AC → Simplest form of propogation makes each arc consistent.

$X \rightarrow Y$ is arc-consistent if & only if every value x of X is consistent with some value y of Y

failure If an assignment lead inconsistency, failure detected

(5)

6. Consider the training examples shown in Table below binary classification problem.

Customer ID	Gender	Car Type	Shirt Size	Class
1	M	Family	Small	C0
2	M	Sports	Medium	C0
3	M	Sports	Medium	C0
4	M	Sports	Large	C0
5	M	Sports	Extra Large	C0
6	M	Sports	Extra Large	C0
7	F	Sports	Small	C0
8	F	Sports	Small	C0
9	F	Sports	Medium	C0
10	F	Luxury	Large	C0
11	M	Family	Large	C1
12	M	Family	Extra Large	C1
13	M	Family	Medium	C1
14	M	Luxury	Extra Large	C1
15	F	Luxury	Small	C1
16	F	Luxury	Small	C1
17	F	Luxury	Medium	C1
18	F	Luxury	Medium	C1
19	F	Luxury	Medium	C1
20	F	Luxury	Large	C1

Figure 3: Decision Tree

(a) Compute the Gini Index for Gender attribute. **(2 Marks)**

Gini (Male) $= 1 - 2 \times (0.5)^2 = 0.5$

Gini (Female) $= 0.5$

0.32 0.16 0.36

Gini (Gender) $= 0.5 \times 0.5 + 0.5 \times 0.5 = 0.5$

(b) Compute the Gini Index for Car Type attribute using multiway split. **(2 Marks)**

Gini (Family) $= 0.375$

Gini (Sports) $= 0$

Gini (Luxury) $= 0.2188$ Overall Gini $= 0.1625$

(c) Compute the Gini Index for Shirt Size attribute using multiway split. **(2 Marks)**

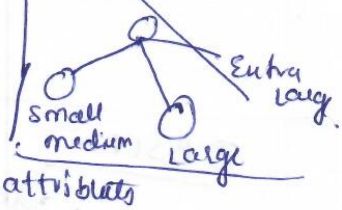
Gini (Small) $= 0.48$

Gini (Medium) $= 0.4898$

Gini (Large) $= 0.5$

Gini (Ex. Large) $= 0.5$ Overall $= 0.4914$

Extra
This might be done like



(d) Which attribute is better, Gender, Car Type, or Shirt Size? **(1 Marks)**

Car Type as it has lowest gini among all 3 attributes

7. Answer the following with respect to clustering:

(a) What is Hierarchical Clustering and Partitional Clustering? **(1 Mark)**

HC → Set of nested clusters organized as hierarchical tree

PC → A division of data objects into non overlapping subsets (clusters) such that data object is in exactly one subset.

(6)

(b) Define a core point and border point in DBSCAN? (**2 Mark**)

CP → A point is a core point if it has atleast a specified number of points (minpts) within Eps

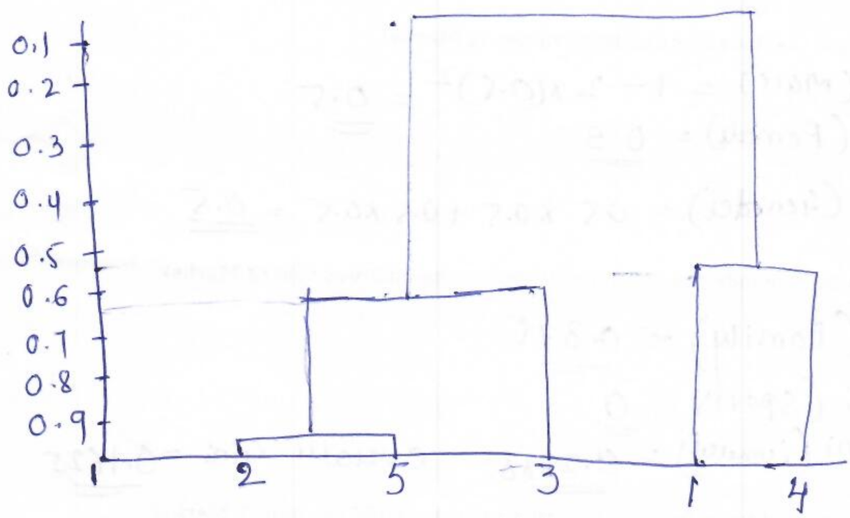
↳ There are points that are at interior of cluster

↳ counts the point itself.

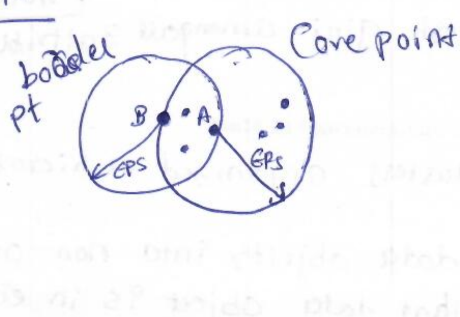
BP → The point which is neighborhood of core point

(c) Use the similarity matrix in Table given below to perform complete link hierarchical clustering. Show your results by drawing a dendrogram. The dendrogram should clearly show the order in which the points are merged. *Hint*: In complete link, proximity of two clusters is based on the two most distant points in the different clusters (**5 Marks**)

	p1	p2	p3	p4	p5
p1	1.00	0.10	0.41	0.55	0.35
p2	0.10	1.00	0.64	0.47	0.98
p3	0.41	0.64	1.00	0.44	0.85
p4	0.55	0.47	0.44	1.00	0.76
p5	0.35	0.98	0.85	0.76	1.00



DBSCAN



Q.7(c)

Sol: In complete link, proximity of two cluster is based on the two most distant points.

	P_1	P_2	P_3	P_4	P_5
P_1	1.00	0.10	0.41	0.55	0.35
P_2	0.10	1.00	0.64	0.47	0.98
P_3	0.41	0.64	1.00	0.44	0.85
P_4	0.55	0.47	0.44	1.00	0.76
P_5	0.35	0.98	0.85	0.76	1.00

The max distant is between P_2 and P_5. So first cluster is formed between P_2 & P_5.

Now we need to recalculate the matrix; so for that we'll update (a/c) to $\min[\text{dist}(P_2,P_5), P_i]$ where $i = 1, 3, 4$. for eg: $\min(\text{dist}(P_2,P_5),P_1) = 0.10$

	P_1	P_3	P_4	P_2P_5
P_1	1.00	0.41	0.55	0.10
P_3	0.41	1.00	0.44	0.64
P_4	0.55	0.44	1.00	0.47
P_2P_5	0.10	0.64	0.47	1.00

Now the max distant is between (P_2, P_5) and P_3.

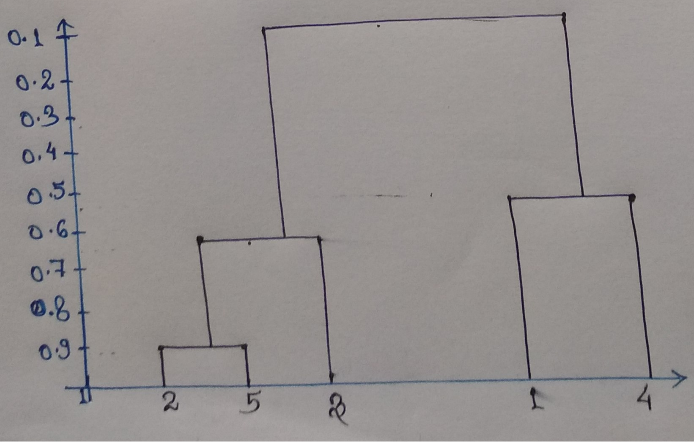
Again we'll use above steps & the updated matrix will be:

	P_1	P_4	$P_2P_5P_3$
P_1	1.00	0.55	0.10
P_4	0.55	1	0.44
$P_2P_5P_3$	0.10	0.44	1.00

Now the max distant is between P_1, and P_4.

	P_1P_4	$P_2P_5P_3$
P_1P_4	1.00	0.10
$P_2P_5P_3$	0.10	1.00

Now at the last max distant is (b/w) (P_1, P_4) and $(P_2P_5P_3)$.



7

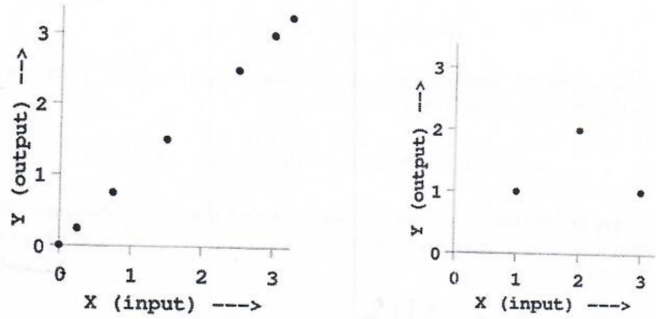
8. Regression

(a) Why do we add a regularization term added in the objective function? (**1 Marks**)

To Avoid overfitting

(b) For the Data Set-1 (from Figure 4)

x	y	(x-y)
1	1	0
2	2	0
3	3	0
4	4	



X	Y
1	1
2	2
3	1

x-y	$(x-y)^2$
0	0
0	0
2	4

(a) Data Set-1 (b) Data Set-2

Figure 4: Data Sets

i. what is mean squared training set error of the running linear regression (using the model $y = w_0 + w_1 x$) (**2 Marks**)

0

ii. what is mean squared test set error of the running linear regression assuming the righmost three points are in the test set and the others are in the training set. (**2 Marks**)

0

(c) For the Data Set-2 (from Figure 4)

i. what is mean squared training set error of the running linear regression (using the model $y = w_0 + w_1 x$) (**2 Marks**)

MSE = 2/9

9. Support Vector Machines

8

(a) One of the most commonly used kernels in SVM is the Gaussian RBF kernel: $k(x_i, x_j) = exp(-\frac{||x_i - x_j||^2}{2\sigma})$. Suppose we have three points, z_1, z_2, and x. z_1 is geometrically very close to x, and z_2 is geometrically far away from x. What is the value of $k(z_1, x)$ and $k(z_2, x)$?. Circle one of the following: **(1 Mark)**

i. $k(z_1, x)$ will be close to 1 and $k(z_2, x)$ will be close to 0.
ii. $k(z_1, x)$ will be close to 0 and $k(z_2, x)$ will be close to 1.
iii. $k(z1, x)$ will be close to c_1, $c_1 \gg 1$ and $k(z2, x)$ will be close to c_2, $c_2 \ll 0$, where $c_1, c_2 \in R$
iv. $k(z_1, x)$ will be close to c_1, $c_1 \ll 0$ and $k(z_2, x)$ will be close to c_2, $c_2 \gg 1$, where $c_1, c_2 \in R$

(b) Kernel functions implicitly define some mapping function $\phi(.)$ that transforms an input instance $x \in R^d$ to a high dimensional feature space Q by giving the form of dot product in

$$Q : K(x_i, x_j) = \phi(x_i).\phi(x_j).$$

Assume we use radial basis kernel function $K(x_i, x_j) = exp(-\frac{1}{2}||x_i - x_j||^2)$. Thus we assume that there's some implicit unknown function $\phi(x)$ such that

$$\phi(x_i).\phi(x_j) = K(x_i, x_j) = exp(-\frac{1}{2}||x_i - x_j||^2).$$

Prove that for any two input instances x_i and x_j, the squared Euclidean distance of their corresponding points in the feature space Q is less than 2, i.e. prove that $||\phi(x_i) - \phi(x_j)||^2 < 2$. **(3 marks)**

$$||\phi(x_i) - \phi(x_j)||^2$$
$$= [\phi(x_i) - \phi(x_j)) \cdot (\phi(x_i) - \phi(x_j))$$
$$= \phi(x_i) \cdot \phi(x_i) + \phi(x_j) - 2 \cdot \phi(x_i) \cdot \phi(x_j)$$
$$= 2 - 2exp(-\frac{1}{2}||x_i - x_j||^2)$$
$$< 2$$

10. How do we address the issue of zero conditional probabilities in Naive Bayes classification? **(2 marks)**

$$P(x_i = c|y) = \frac{nc + mp}{n + m}$$

if n, m, P, v explained.

Laplace estimate

$$P(x_i = c|y) = \frac{nc + 1}{n + v}$$

*******oOo*******