Name: Roll number:

CS 302 – Artificial Intelligence | Endsem Exam | March 23, 2022

Maximum marks – 75 | Weight – 40 %. | *Be Concise*

1. **Naïve Byes:** Consider the data set shown in Figure 1 below **[5*2= 10]**
   a. Estimate the conditional probabilities for P(A|+), P(B|+), P(C|+), P(A|−), P(B|−), and P(C|−).
   b. Use the estimate of conditional probabilities given in the previous question to predict the class label for a test sample (A = 0, B = 1, C = 0) using the naive Bayes approach.
   c. Estimate the conditional probabilities using the m-estimate approach, with p = 1/2 and m = 4.
   d. Repeat part (b) using the conditional probabilities given in part (c).
   e. Compare the two methods for estimating probabilities. Which method is better and why?

Record	A	B	C	Class
1	0	0	0	+
2	0	0	1	−
3	0	1	1	−
4	0	1	1	−
5	0	0	1	+
6	1	0	1	+
7	1	0	1	−
8	1	0	1	−
9	1	1	1	+
10	1	0	1	+

Figure 1

2. **Clustering:** Answer related to clustering **[2+2+6=10]**
   a. What is single link and Completer link in hierarchical clustering?
   b. What are core point and border point?
   c. This question is on k-means clustering to cluster the points A, B . . . F (indicated by ○ in the **figure 2**) into 2 clusters. The current cluster centers are P and Q (indicated by the ■ in the diagram on the right). Recall that k-means requires a distance function. Given 2 points, A = (A1, A2) and B = (B1, B2), we use the following distance function d(A, B) that you saw from class,

$$d\,(A, B) = (A_1 - B_1)^2 + (A_2 - B_2)^2$$

      i. Update assignment step: Select all points that get assigned to the cluster with center at P:
      ii. Update cluster center step: What does cluster center P get updated to?

   Now consider the Manhattan distance:

$$d'\,(A, B) = |A_1 - B_1| + |A_2 - B_2|$$

We again start from the original locations for P and Q as shown in the figure, and do the update assignment step and the update cluster center step using Manhattan distance as the distance function:

      iii. Update assignment step: Select all points that get assigned to the cluster with center at P, under this new distance function d' (A, B).

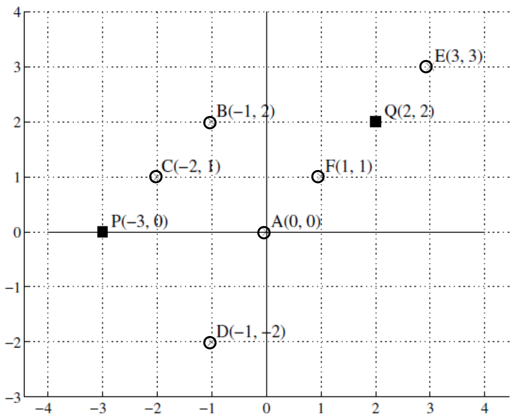


Figure 2

3. **Decision Trees**: Answer the questions related to the data set given below (**Figure 3**). You will try to find whether Nitesh finds a particular type of food Appealing based on the food's temperature, taste, and size.

[2+2+4=8]

Appealing	Temperature	Taste	Size
No	Hot	Salty	Small
No	Cold	Sweet	Large
No	Cold	Sweet	Large
Yes	Cold	Sour	Small
Yes	H	Sour	Small
No	H	Salty	Large
Yes	H	Sour	Large
Yes	Cold	Sweet	Small
Yes	Cold	Sweet	Small
No	H	Salty	Large

Figure 3

a. What is the initial entropy of Appealing?
b. Let Taste be chosen for the root of the decision tree. What is the information gain associated with this attribute?
c. Draw the full decision tree learned for this data.

4. **Search:** Answer the following questions for the graph below (**Figure 4**). Break any ties alphabetically. For the questions that ask for a path, please give your answers in the form 'S – A – D – G.' Here, S is the Start node and G is the Goal node.

[2+2+2+2+4=12]

a. What path would breadth-first graph search return for this search problem?
b. What path would uniform cost graph search return for this search problem?
c. What path would depth-first graph search return for this search problem?
d. What path would A* graph search, using a consistent heuristic, return for this search problem?

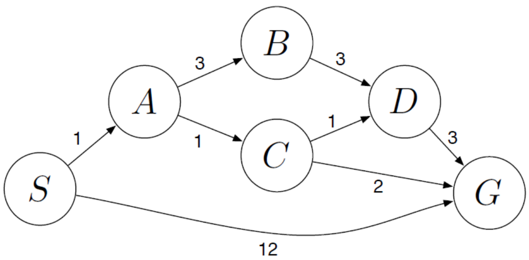


Figure 4

e. For the heuristics for this problem shown in the table below (**Figure 5**), answer Yes or No. No need to give explanation.

State	h_1	h_2
S	5	4
A	3	2
B	6	6
C	2	1
D	3	3
G	0	0

Figure 5

(i) h1 is admissible? Yes, or No
(ii) h1 is consistent? Yes, or No
(iii) h2 is admissible? Yes, or No
(iv) h2 is consistent? Yes, or No

Roll Number______________________

5. **SVM and Feature space:** A kernal function K(x, z) is a function gives the similarity between two instances x and z in a transformed space. i.e., for a feature transform x → ϕ(x), the kernel function is K(x, z) = ϕ(x) · ϕ(z). Let us explore some kernel functions and their feature transforms. The input vectors are in 2 dimensional (i.e. x = (x1, x2)). Remember that x · z = x1z1 + x2z2. **[6+6=12]**

   a. Mention the feature transform for each of the kernel functions below, (mention a single option only for each question)

      (i) $K(x, z) = 1 + x \cdot z$

      - ○ $\phi(x) = (x_1, x_2)$
      - ○ $\phi(x) = (x_1^2, x_2^2)$
      - ○ $\phi(x) = (1, x_1, x_2)$
      - ○ $\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$
      - ○ $\phi(x) = (1, x_1^2, x_2^2)$
      - ○ $\phi(x) = (1, x_1^2, x_2^2, \sqrt{2}x_1x_2)$

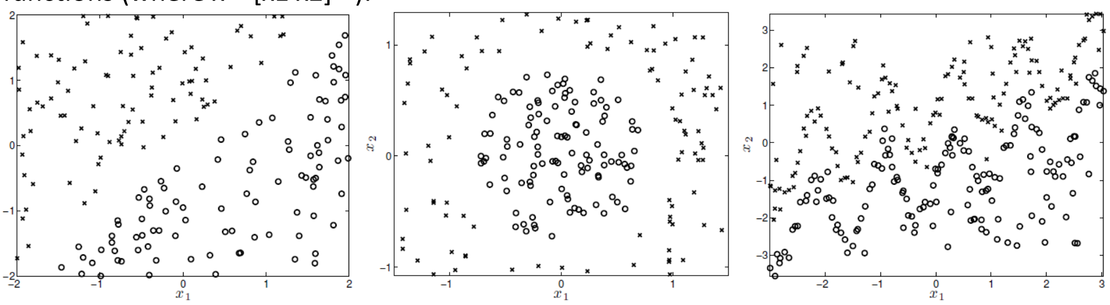
      (ii) $K(x, z) = (x \cdot z)^2$

      - ○ $\phi(x) = (x_1^2, x_2^2)$
      - ○ $\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$
      - ○ $\phi(x) = (1, x_1^2, x_2^2)$
      - ○ $\phi(x) = (1, x_1, x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$
      - ○ $\phi(x) = (1, x_1^2, x_2^2, \sqrt{2}x_1x_2)$
      - ○ $\phi(x) = (x_1, x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$

      (iii) $K(x, z) = (1 + x \cdot z)^2$

      - ○ $\phi(x) = (1, x_1^2, x_2^2)$
      - ○ $\phi(x) = (1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2)$
      - ○ $\phi(x) = (1, x_1^2, x_2^2, \sqrt{2}x_1x_2)$
      - ○ $\phi(x) = (1, x_1, x_2, \sqrt{2}x_1x_2)$
      - ○ $\phi(x) = (1, x_1^2, x_2^2, x_1, x_2, \sqrt{2}x_1x_2)$
      - ○ $\phi(x) = (1, x_1x_2, x_1^2x_2^2)$

   b. The pictures below (**Figure 6**) represent three distinct two-dimensional datasets with positive examples labeled as o's and negative examples labeled as x's. Consider the following three kernel functions (where x = [x1 x2]ᵀ):



(A) (B) (C)

Figure 6

      (i) Linear kernel: $K(x, z) = x^\top z = x \cdot z = x_1z_1 + x_2z_2$

      (ii) Polynomial kernel of degree 2: $K(x, z) = (1 + x^\top z)^2 = (1 + x \cdot z)^2$

      (iii) RBF (Gaussian) kernel: $K(x, z) = \exp\left(-\frac{1}{2\sigma^2}\|x - z\|^2\right) = \exp\left(-\frac{1}{2\sigma^2}(x - z)^\top (x - z)\right)$

   For each dataset (A, B, C) mention all kernels that make the dataset separable (assume σ = 0.01 for the RBF kernel)

6. **CSP** **[4+2=6]**

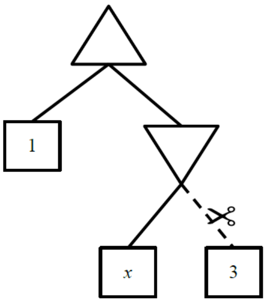
   a. Soham (S), Bhuvan (B), Mohan (M), Dhruv (D), Pavan (P), and Gaurav (G) are lining up next to each other. The positions are numbered 1, 2, 3, 4, 5, 6, where 1 neighbors 2, 2 neighbors 1 and 3, 3 neighbors 2 and 4, 4 neighbors 3 and 5, 5 neighbors 4 and 6, and 6 neighbors 5. Each one of them takes up exactly one spot. Bhuvan (B) needs to be next to Mohan (M) on one side and Dhruv (D) on the other side. Pavan (P) needs to be next to the Gaurav (G). Soham (S) needs to be at 1 or 2. Formulate this problem as a CSP: list the variables, their domains, and the constraints. Encode unary constraints as a constraint rather than pruning the domain. (No need to solve the problem, just provide variables, domains and implicit constraints.)

      - Variables
      - Domain
      - Constraints

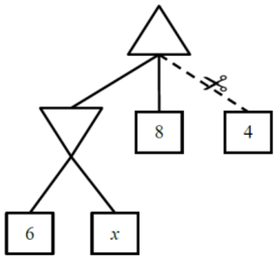
b. Consider a CSP with variables A, B with domains {1, 2, 3, 4, 5, 6} for A and {2, 4, 6} for B, and constraints A < B and A + B > 8. List the values that will remain in the domain of A after enforcing arc consistency for the arc A → B (recall arc consistency for a specific arc only prunes the domain of the tail variable, in this case A).

7. For each of the game-trees shown below, state for which values of x the dashed branch with the scissors will be pruned. If the pruning will not happen for any value of x write "none". If pruning will happen for all values of x write "all", if pruning will happen for a range please specify the with ≤ x or ≥ x.

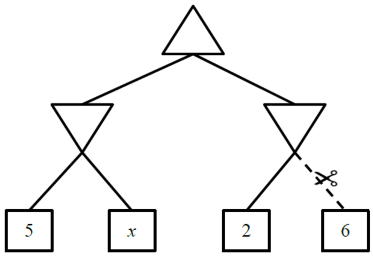
[2+2+2+3+4=12]



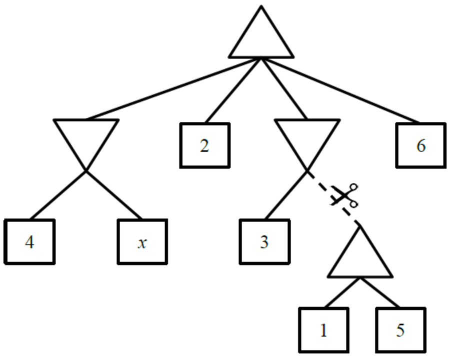
a)



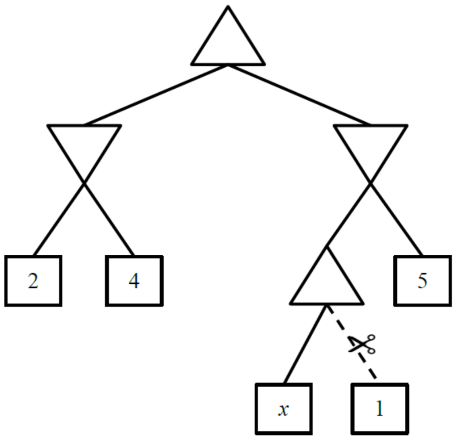
b)



c)



d)



e)

8. Regression **[2+3=5]**

a. What is Overfitting and how can it be reduced?

b. What is the role of regularization parameter λ? What happens if λ =0 and λ →∞ in the cost function.

----------oOo--------