Question Paper

Exam Date & Time: 09-Jan-2024 (02:30 PM - 05:30 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

FIFTH SEMESTER B.TECH END SEMESTER MAKE UP EXAMINATIONS, JAN 2024

MACHINE LEARNING [CSE 3171]

Marks: 50 **Duration: 180 mins.**

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1) Compare and contrast the supervised with unsupervised learning. (4)

A)

B) Discuss the importance of formulating well-posed machine learning problems in the development of intelligent machines. Provide a real-world example illustrating the impact of problem formulation on the success of a machine learning application. (3)

C) Discuss the significance of Occam's Razor Principle and Overfitting Avoidance in the context of machine learning. Provide an in-depth analysis of how these principles contribute to model selection and generalization. Illustrate with examples where the application of these principles has influenced real-world machine learning systems. (3)

2) Summarize the Pros and Cons of K-Nearest Neighbors (KNN). (4)

A)

B) Suppose you are given the following set of data listed in Table 2B with three Boolean input variables a, b, and c, and a single Boolean output variable K. (3)

Table 2B

a	b	c	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

For parts (a) and (b), assume we are using a naive Bayes classifier to predict the value of K from the values of the other variables.

a) According to the naive Bayes classifier, what is P(K = 1|a = 1 ∧ b = 1 ∧ c = 0)?

b) According to the naive Bayes classifier, what is P(K = 0|a = 1 ∧ b = 1)?

C) Consider the following Figure 2C Bayesian network. Assume that: (3)

P(Alarm1) = 0.1

P(Alarm2) = 0.2

P(Burglary | Alarm1, Alarm2) = 0.8

P(Burglary | Alarm1, ¬ Alarm2) = 0.7

P(Burglary | ¬ Alarm1, Alarm2) = 0.6

P(Burglary | ¬ Alarm1, ¬ Alarm2) = 0.5

Calculate P(Alarm2 | Burglary, Alarm1). Show all of your reasoning.

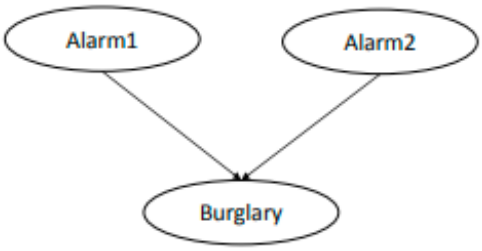

Fig. 2C

3) Summarize the main advantages and limitations of SVM. (4)

A)

B) The data listed in Table 3B, show the sugar content of a fruit (SUGAR) for different numbers of days after picking (DAYS). (3)

Table 3B

Days	Sugar
0	7.9
1	12.0
3	9.5
4	11.3
5	11.8
6	11.3
7	4.2
8	0.4

Obtain the estimated regression line to predict sugar content based on the number of days the fruit is left on the tree.

C) Consider a dataset shown in Table 3C below with two features, X1 and X2, and two classes, labelled 0 and 1. (3)

Table 3C

X1	X2	Class
2	3	0
1	4	1
3	2	0
4	5	1
2	1	0

Train an SVM model with a linear kernel on this dataset. Assume the following support vectors and their corresponding coefficients:

1. Support Vector 1: (2, 3) with coefficient $\alpha_1 = 0.5$
2. Support Vector 2: (1, 4) with coefficient $\alpha_2 = 0.3$

Calculate the decision function for a new data point (3, 2). Determine the class prediction for the data point (3, 2) based on the decision function.

4) Use single link agglomerative hierarchical clustering to group the data described as {18, 22, 25, 27}. Clearly show the proximity matrix corresponding to each iteration of the algorithm and plot the dendrogram. Discuss on the limitations of agglomerative clustering technique. (4)

A)

B) Describe two cases where K-Means clustering fails to give good results. Show pictorially a two-dimensional data space where K-Means cluster analysis fails. (3)

C) Generate Confusion matrix for an image recognition problem having actual values = {1,1, 0, 1, 0, 0, 1, 0, 0, 0} and predicted values = {1, 0, 0, 1, 0, 0, 1, 1, 1, 0}. Derive classification metrics such as accuracy, precision, recall and F1-Score. (3)

5) With an example data, work out a case where the entropy calculation would be zero. Relating zero entropy with information gain, explain two scenarios which consider information gain as undesirable metric. (4)

A)

B) Why decision tree is considered as a nonparametric model? In comparison to K- NN as a nonparametric model, give a scenario where decision tree is expected to have higher performance than K-NN. (3)

C) Why do overfitting occur in decision trees? Give a technique to solve overfitting and illustrate the same with an example tree. (3)

-----End-----