Introduction to Machine Learning

GCN/SV/KPS BTECH-V SEM

24, November 2021 Time 11:00 AM- 2:00 PM

Question 01
Consider the following binary classification dataset for the prediction of label K given the features a, b and c.

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

1. According to the naive Bayes classifier, what is $P(K = 1|a = 1 \ \Lambda \ b = 1 \ \Lambda \ c = 0)$.
2. Estimate $P(K = 1|a = 0 \ \Lambda \ b = 0 \ \Lambda \ c = 1)$.
3. Investigate if the above prediction(s) suffer from the zero-frequency problem; if they do, suggest a corrective course of action before making the predictions. Any assumptions taken must be specified before making the corrected predictions.

Question 02

Explain the utility of universal approximation theorem in context to artificial neural networks. Apply this theorem to approximate $f(\mathbf{x}) = e^{\mathbf{x}}$ in the following setting:

1. Assume that the input $\mathbf{x} \in \Re^3$.
2. The artificial neural network contains two hidden layers (each having 2 hidden units) in addition to the input and the output layer.
3. For simplicity, assume that the activation in the hidden layers is linear given by $a\mathbf{x} + b$.

Clearly show the weight updates in each node for the first iteration. You may select your own values for the input $\mathbf{x}$ and hyperparameters a and b. Any other assumption(s) if taken taken must be mentioned in the answers.

Question 03 [25 Marks]

Compute the PCs for the data set (1,2), (2,3), (3.4), (4,5), (5,6), (6,8). Show all the intermediate steps in detail.

Question 04 [25 Marks]

Build a DT of depth two for the data set ((4,1;1).(6,6;0), (9,5;1), (1,2;0), (7,3;1), (5,4;0)). The first two entries are attributes (A1 and A2) and the third is the binary label.

1. Give the rules intuitively for the first and second split without resorting to entropy calculation? Write your answer in a form like $A1 > a$ or $A2 > b$.
2. Suppose we are trying to build the DT from scratch. Is there a nontrivial split at the root that would give an information gain of zero?

Question 05 [3+3+8+6 Marks]

What is the speciality of Support Vector Machine? Who first invented it and when? Formulate the primal problem of SVM. Following are the dataset having two features: (1,1); (1, -1); (0.5, 0.75) ;(0.5, -0.75); (3,1) ;(3, -1)- belong to class-1 and (4,0) ;(5,1) ;(6,0) ;(5, -1);(7,2)- belong to the class-2. Using SVM, find the equation of the separating hyperplane.

Question 06 [5+5+5+5 Marks]

With illustrative examples, explain the concept of Lagrangian duality? Using indicator function, prove that in general the dual problem, $D^* \leq P^*$ (primal problem). Formulate SVM dual problem mentioning why SVM dual problem is required to be formulated. What is KKT condition and what is its relevance in formulating SVM dual problem?

Question 07 [5+10+5 Marks]

With example, explain the concept of soft margin support vector machine. Reformulate the dual problem of SVM using regularization. What will be the performance of a SVM classifier, if you have noisy data?

Question 08 [5+5+10 Marks]

What is Co-ordinate ascent algorithm? Can it be used for training a SVM? Explain the basic working principle of SMO algorithm.

Question 09 [5+5+5+5 Marks]

How a SVM can be used for nonlinear classification task? Explain the concept of Kernel trick with suitable examples. Name 4-5 kernel functions. How to design your own kernel?