

CE1.2-R4 : MACHINE LEARNING

NOTE :

1. Answer question 1 and any FOUR questions from 2 to 7.
2. Parts of the same question should be answered together and in the same sequence.

Time : 3 Hours

Total Marks : 100

1. (a) What are learning systems in context of machine learning ? What are the steps involved in designing a learning system ?
(b) What is the use of hard margin and soft Margin in SVMs ? Explain with a diagram.
(c) Discuss, how recurrent neural networks are trained ?
(d) How does bayes theorem works in terms of prior, evidence and likelihood ? What is the best data set scenario for the Naïve Bayes Classifier ?
(e) What are the Advantages of Converting Decision Tree to Rules Before Pruning ?
(f) What is Overfitting Avoidance Bias ?
(g) List any two real-time applications of SVM on classification and regression problems. (7x4)
2. (a) Write and explain the algorithm for Candidate-Elimination Method.
(b) What are the Limitations of the Find-S Algorithm ?
(c) Can SVM be used for regression ? If yes, then describe how Support Vector Machine is different in Classification and Regression. (6+6+6)
3. (a) What is a confusion matrix ? How various performance measures are computed using a confusion matrix ? Explain with an example.
(b) How are ensemble methods in machine learning provide better than solitary models of machine learning in remote sensing ?
(c) Explain sensitivity and specificity. (8+6+4)
4. (a) What is the significance of statistical hypothesis testing ? How a statistical hypothesis is performed ?
(b) What are the benefits of cross-validation ? Explain the process of k-fold cross-validation with a suitable example.
(c) What are various prioritization score approaches used for active learning ? Explain in brief. (6+8+4)

5. (a) What is Laplace smoothing and why do we need it in a Naive Bayes classifier ? Explain with example.
- (b) Briefly explain various functions associated with the Kernel function along with equations and graphs.
- (c) Justify the importance of kernel and Decision Boundary in SVM. How does a large value of the C Parameter affect SVM ?

(4+8+6)

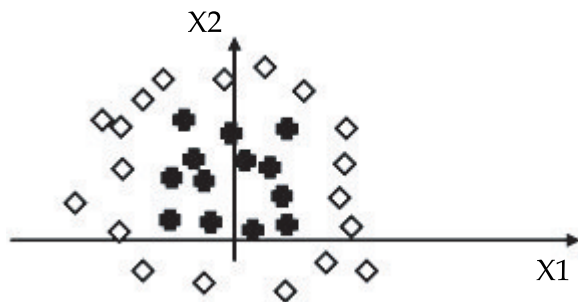
6. (a) Suppose a new test is developed by a medical scientist to detect a disease which gives almost accurate results. The accuracy of the test is 99%, that is, with probability 0.99 it gives the correct result (the same probability for disease-positive-test and no-disease-negative-test combinations) and only in 1% of tested cases (probability 0.01) the result is wrong. The incidence of the disease in the population is 0.01% (probability 0.0001). Compute the probability that somebody from wide population who has tested positive indeed suffers from the disease.

Would you recommend the test to be widely adopted ?

- (b) What is regularization in Logistic regression and what are its different types ? When do we use which type of regularization in Logistic regression ?
- (c) What does training a logistic regression model mean ?

(10+6+2)

7. (a) What is the use of Function Approximation in machine learning ?
- (b) What is the importance of 'training Set' and 'test Set' in a Machine Learning Model ? How allocation is performed between Training Validation, and Test Sets ? Discuss with examples.
- (c) Given a set of training data which is plotted as below :



Build a neural network that classifies two dimensional data (i.e., $X = [x_1, x_2]$) into two classes. Justify your choice of the number of nodes and the architecture.

Draw the decision boundary that your network can find on the diagram.

(4+8+6)

- o O o -