011705

December–2023

B.Tech (IT) 7th Semester

Basics of Machine Learning (PEC-CS-D-501)

Time : 3 Hours]   [Max. Marks : 75

Instructions :

1. *It is compulsory to answer all the questions (1.5 marks each) of Part-A in short.*
2. *Answer any four questions from Part-B in detail.*
3. *Different sub-parts of a question are to be attempted adjacent to each other.*

PART–A

1. (a) What is hypothesis, hypothesis space, feature, feature space? (1.5)
   (b) What is underfitting and overfitting and trade-off between them'? (1.5)
   (c) What is k-fold Cross validation technique? (1.5)
   (d) Describe regression, classification, supervised and unsupervised learning. (1.5)
   (e) Differentiate between Training data and Testing and validation data. (1.5)

(f) Define (i) Independents and Dependent probabilities (ii) Conditional Probability (iii) Prior and Posterior Probability (1.5)

(g) How entropy and information gain are related. (1.5)

(h) What is Artificial Neural Network? (1.5)

(i) Explain the tradeoff between bias and variancea. (1.5)

(j) Why Näive bayes' is called Naïve? (1.5)

PART–B

2. (a) Discuss Linear Regression with an example. Derive formula for MLE cost function. (10)

(b) State the gradient updates for both batch gradient descent and stochastic gradient descent. (5)

3. (a) Describe K-Nearest Neighbour method. Explain with its pros and cons. (5)

(b) What is confusion matrix? What are true positive, true negative, false positivc, false negative with an example? Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the precision and recall for the data. (10)

4. State the mathematical formulation of the SVM as an optimization problem. Describe the significance of soft margin hyperplane and explain how they are computed. Specify the significance of Kernal functions in SVM. Describe any two Kernel functions. (15)

5. (a) List the issues in Decision Tree Learning and specify the possible solutions. (5)

(b) What is an ensemble? Explain bagging, boosting and random forest ensemble techniques. (10)

6. (a) What are various components of Time series? What do you mean by random walk? (7.5)

(b) What is Sparse Modelling? Explain how sparsity of a model be used to improve computational efficiency during training? (7.5)

7. Explain PCA with calculation of Eigen vectors and derivation of covariance. (15)