2019

COMPUTER SCIENCE AND ENGINEERING

Paper : CSEL 0919

(Elective-II : Machine Learning)

Full Marks : 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

1. Answer ***any five*** from the following : 2×5

(a) What is machine learning? Explain with an example.

(b) What are supervised learning and unsupervised learning?

(c) Differentiate between regression and classification.

(d) Formulate the cost function of logistic regression.

(e) What is logistic function?

(f) What is bias variance trade-off? What are its implications?

(g) What is a perceptron?

2. Answer ***any five*** from the following : 4×5

(a) What are commonly used metrics to assess the performance of classification modules?

(b) What is a polynomial regression? How it can be represented in the form of a matrix?

(c) Suppose the classifier's prediction is given as follows :

		Predicted	
Actual		+	–
	+	60	15
	–	10	15

Calculate Accuracy, Weighted Accuracy, Precision and Recall for it.

(d) What is cross validation? How it improves the accuracy of the outcomes?

(e) Define and explain Squared Error (SE) and Mean Squared Error (MSE) w.r.t. to Regression.

(f) Explain Geometric models and Probabilistic models of machine learning in detail.

(g) Critically comment on overfitting and underfitting.

Please Turn Over

3. Answer ***any four*** from the following :

(a) (i) A dealer has a warehouse that stores a variety of fruits and vegetables. When fruit is brought to the warehouse, various types of fruit may be mixed together. The dealer wants a model that will sort the fruit according to type. Justify with reasons how efficient is machine learning model compared to feature based classification technique.

(ii) Why it is necessary to estimate the accuracy of hypothesis? Explain procedure to estimate difference in error between two learning methods. 5+5

(b) What is a recommender system? What is the difference with query based searching? What is a recommendation engine? How does it work? What is collaborative filtering? (2+1+4+3)

(c) (i) What is a Support Vector? How the margins are defined in support vector machine?

(ii) Explain regression using least square method. What is multiple regression? 5+5

(d) What is PCA? Mathematically derive 1st & 2nd Principal Components. Why is normalization of variables necessary? What is the appropriate matrix (covariance or correlation) in principal component analysis? 1+4+2+3

(e) What is a K-nearest neighbor classifier? Write down the KNN algorithm pseudocode. What are the advantages and limitations of this algorithm? What are different types of KNN algorithms? Comment on the performance of the algorithm on different data volumes. 1+3+3+1+2

(f) (i) How to show or prove a dataset is not linearly separable?

(ii) How are the weights updated in the perceptron learning rule?

(iii) What is gradient descent and delta rule? Why stochastic approximation to gradient descent is needed? 3+3+4
