Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 70401

M.E./M.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2019

First Semester

Biometrics and Cyber Security

CP 5191 – MACHINE LEARNING TECHNIQUES

(Common to M.E. Computer Science and Engineering/M.E. Computer Science and Engineering (With Specialization in Networks)/M.E. Mobile and Pervasive Computing/M.E. Software Engineering/M.Tech. Information Technology)

(Regulations 2017)

Time : Three Hours — Maximum : 100 Marks

Answer ALL questions

PART – A **(10×2=20 Marks)**

1. State the different types of machine learning algorithms.
2. When to use regression ?
3. Discuss the advantages of SVM.
4. What is an RBF network ?
5. Write the Adaboost algorithm for boosting.
6. How does ensemble of classifiers improve the accuracy of a classification system ?
7. Distinguish model based and model free learning.
8. Justify the necessity for dimensionality reduction in the context of machine learning.
9. Construct a Bayesian network simulating a teacher entering and leaving a class.
10. Where Markov Random fields will be useful ?

PART – B **(5×13=65 Marks)**

11. a) i) Discuss the candidate elimination algorithm and suggest alternate hypothesis for the same. **(8)**

ii) Explain the steps in designing a learning system. **(5)**

(OR)

b) i) What is a neuron ? Map it with an artificial neuron. (5)

ii) A company manufactures an electronic device to be used in a very wide temperature range. The company knows that increased temperature shortens the life time of the device and a study is therefore performed in which the life time is determined as a function of temperature. Express the life time as a linear function of the temperature. The following data is found : (8)

Temperature (C)	10	20	30	40	50	60	70	80	90
Lifetime (Hours)	420	365	285	220	176	117	69	34	5

12. a) Draw the model and explain the algorithm for back propagation. Derive necessary equations to depict the back propagation error. (13)

(OR)

b) Explain the Support vector machine from the perspective of a non-linear kernel by means of an algorithm. Derive the Margin of the support vectors with an example and depict it with necessary diagrams. (13)

13. a) Use the k-means algorithm to cluster the following 8 points into 3 clusters: A1 = (2,10), A2 = (2,5), A3 = (8,4), A4 = (5,8), A5 = (7,5), A6 = (6,4), A7 = (1,2), A8 = (4,9). Suppose that the initial seeds (centers of each cluster) are Al, A4 and A7. Run the k-means algorithm for 2 epochs. At the end of this epoch show :

i) The new clusters (i.e. the points belonging to each cluster); (3)

ii) The centers of the new clusters; (3)

iii) Draw a 10 by 10 space with all the 8 points and show the clusters after the second epoch and the new centroids. (3)

iv) How many more iterations are needed to converge ? Draw the result for each epoch. (4)

(OR)

b) i) Construct a decision tree for the expression A = X AND Y OR Z. (8)

ii) Explain GMM with a neat diagram. (5)

14. a) Explain the PCA algorithm with necessary examples. (13)

(OR)

b) Discuss Genetic algorithm with an example. Explain mutation, crossover, chromosomes and generations. (13)

15. a) Write the forward algorithm and Viterbi decoder for a HMM. (13)

(OR)

b) Explain Inference in Bayesian networks with an example. (13)

PART – C (1×15=15 Marks)

16. a) You are asked to solve the problem of identifying fake bloggers. A fake blogger is one who has posted at least 10% of fake blogs. As you are aware, blogs can be in multiple languages and individual blogs can be in a mix of language. You have been given the identifiers of nearly 1000 bloggers and you have been told that at least 20% of the bloggers post fake blogs. You decided to solve the problem using a machine learning approach. Answer the following questions in this context :

a) What are the features that you would consider ? Name them and justify their relevance. (4)

b) What approach do you use to extract these features ? Discuss how do you store them. (4)

c) Determine the appropriate machine learning algorithm and construct a flow diagram depicting how you would determine the fake bloggers. (5)

d) Discuss the metrics used to verify your algorithm. (2)

(OR)

b) You are asked to be a judge for a Mr. X and Ms. Y contest. You decided to automate the process of awarding Mr. X and Ms. Y by applying an appropriate machine learning algorithm as the number of candidates is more than 1000 for each Mr. X and Ms. Y. Answer the following questions for this scenario.

a) What are the features that you would consider ? Justify. (4)

b) Determine the machine learning approach - supervised, unsupervised or semisupervised. (2)

c) What approach do you use to extract these features ? (2)

d) Determine the appropriate machine learning algorithm and construct a flow diagram depicting how you would determine Mr. X and Ms. Y. (5)

e) Discuss the metrics used to verify your algorithm. (2)

———————