III Semester M.C.A. (Two Years Course) Examination, May/June 2025
(CBCS) (2021 – 22)
COMPUTER SCIENCE
MCA 3E1/MSC 3E1 : Machine Learning (Elective)

Time : 3 Hours | Max. Marks : 70

PART – A

Answer **any 5** questions **each** question carries **6** marks. **(5×6=30)**

1. In a test set with 100 elements, 70 elements are cats and 30 elements are non-cats. The machine recognised 60 among 70 cats as cats and 15 among non-cats as cats. Construct confusion matrix. Also find out the accuracy, recall and sensitivity of the test data.
2. What are the applications of machine learning ? What is supervised learning and unsupervised learning ? Give examples.
3. For the following data with two features and given the labelling, find out which label does (4, 1) belong to using 3-nearest neighbour algorithm :

Element	Feature x_1	Feature x_2	Label
1	1	2	1
2	1	3	1
3	4	5	2
4	5	4	2
5	4	4	2

4. What is principal component analysis ? Why are covariance matrix, eigen values and eigen vectors important for finding PCA.
5. What is SVM ? How can one improve the classification using SVM ? Illustrate with your example.
6. Why is random forest algorithm a better algorithm ? Given a data set with 100 elements and 4 features, explain how random forest can be generated ? How are decisions made using these random forests ?
7. Discuss genetic algorithm for classification problem. How do they differ from other classification algorithms ?
8. What are HMMs used for in machine learning applications ?

P.T.O.

PART – B

Answer **any 4** questions **each** question carries **10** marks. **(4×10=40)**

9. Given the following data with two features, cluster the data set using 3-mean clustering algorithm [Assume the centroids are initially at (2, 2), (4, 2), (6, 4)].

Data Set	Feature 1	Feature 2
1	1	2
2	1	4
3	1	6
4	3	4
5	3	6
6	5	1
7	4	4
8	6	2
9	5	6

10. What is Neural network used for ? What is activation function, weight, neuron, bias in neural network ? Given one input layer, two hidden layers and one output layer, give equations for back propagation. Why is back propagation important in neural networking ?

11. What is hierarchical clustering ? Given the following data set with two features, cluster them using hierarchical clustering.

Data Set	Feature 1	Feature 2
1	2	4
2	5	6
3	4	3
4	1	2
5	3	3
6	4	5
7	4	2

12. What is Naive Bayes Algorithm ? Define conditional probability, prior probability. Given 100 mails with 75 being non-spam and 25 being spam mails and words being used as "buy" and "discount" in the mails. Considering suitable probability, find out how can we find the probability of mails being spam given the word as "buy" and to find the probability of mails being non-spam given the words as "buy" and "discount".

13. What is associate rule mining ? Given the following transactions, find the association.

Transaction	Onion	Potato	Burger	Milk	Bread
t_1	1	1	1	0	0
t_2	0	1	1	1	0
t_3	0	0	0	1	1
t_4	1	1	0	1	0
t_5	1	1	1	0	0
t_6	1	1	1	1	1

(Note : 1 for buying; 0 for non-buying)

14. Discuss the following :

a) Decision tree.

b) FP-tree.
