Question Paper

Exam Date & Time: 15-Feb-2021 (10:00 AM - 01:15 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

MANIPAL SCHOOL OF INFORMATION SCIENCES, MANIPAL
FIRST SEMESTER MASTER OF ENGINEERING - ME (MACHINE LEARNING) DEGREE EXAMINATION - FEBRUARY 2021

Applied Machine Learning [MCL 605]

Marks: 100 **Duration: 180 mins.**

MONDAY, FEBRUARY 15, 2021

Answer all the questions.

1) Explain the difference between supervised and unsupervised machine learning? When should you use supervised over unsupervised machine learning? (L2) (TLO no 1.1) (5 Marks) (5)

a.)

b.) Distinguish between classification and regression? When should you use classification over regression? (L2) (TLO no 1.1) (5 Marks) (5)

2) Explain over-fitting and under-fitting with suitable examples. (4 Marks)(L2) (TLO 1.2) (4)

a.)

b.) Write the purpose of k-fold cross-validation techniques. Describe the k-fold cross-validation techniques with suitable example (6 Marks) L3 (TLO 1.2) (6)

3) Employ Gini index to obtain a decision tree using the data set given in the table. (L3) (TLO no 1.3) (10 Marks) (10)

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

4) Explain about 'training Set' and 'test Set' in a Machine Learning Model? How much data will you allocate for the training, validation, and test Sets? (L2) (TLO no 1.3) (4 Marks) (4)

a.)

b.) Relate information gain and Gini impurity splitting criterion used in decision tree approach. (TLO no 1.3) (6 Marks) (L3) (6)

5) Relate bias-variance and overfitting-underfitting with suitable examples (5 marks) L3(TLO 2.1) (5)

a.)

b.) Explain the method of preventing a model from overfitting by using hyperparameter tuning with a suitable example (5 Marks) L2(TLO 2.1) (5)

6) Write the purpose of the confusion matrix and use the confusion matrix given below to calculate the following L3 (TLO 2.1) (10)

- Accuracy (2.5 Marks)
- Misclassification Rate (2.5 Marks)
- Type-1 error (2.5 Marks)
- Type-2 error (2.5 Marks)

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

7) Write the steps by step process of hierarchical clustering. (L3) (TLO 2.2) (3 Mark) (3)

a.)

b.) There are four types of medicines given in the table below. Apply dendrogram approach to identify the number of clusters required in this case and group the medicines into clusters using hierarchical clustering model. (7 Marks) L3(TLO 2.2) (7)

Objects	Attribute 1 (X): weight index	Attribute 2 (Y): pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

8) Compute the Eigenvalues and Eigenvectors from the following data set L3(TLO 2.2) (10 Marks) (10)

x	2.5	0.5	2.2	1.9	3.1	2.3	2	1	1.5	1.1
y	2.4	0.7	2.9	2.2	3	2.7	1.6	1.1	1.6	0.9

9) A set of training examples of the target function and a new instance is presented. The learner is asked to predict the target value or classification for this new instance. Write the steps to derive the target value output by the naive Bayes classifier. (6 Marks) L3(TLO 3.1) (6)

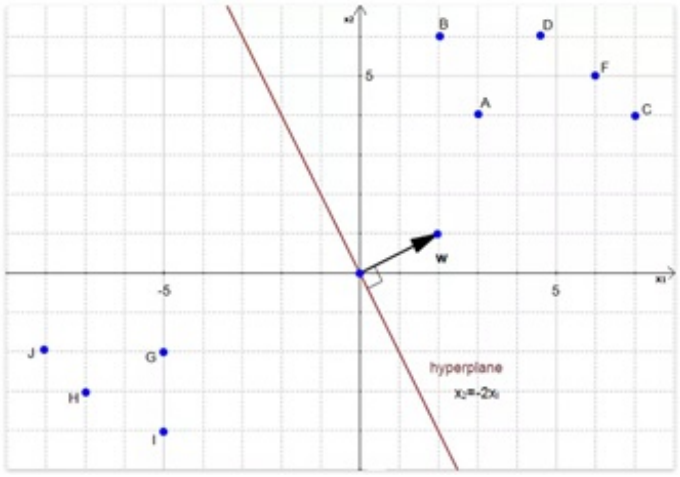
a.)

b.) Define conditional probability. Give an example for the same. (4 marks) L1(TLO 3.1) (4)

10) Compute the following concerning the figure given below: (6)

a.)

- Distance from a point A to the hyperplane and margin of the hyperplane (6 Marks) (TLO 3.2)



b.) Write the principle of Boosting Ensemble models with suitable example (4 Marks) L3(TLO 3.2) (4)

-----End-----