NPTEL Online Certification Courses Indian Institute of Technology Kharagpur

Introduction to Machine Learning

Assignment- Week 1

TYPE OF QUESTION: MCQ

Number of questions: **10** **Total mark: 10 X 2 = 20**

MCQ Question

QUESTION 1:

Which of the following is not a type of supervised learning?

A. Classification
B. Regression
C. Clustering
D. None of the above

Correct Answer: C. Clustering

Detailed Solution : Classification and Regression are both supervised learning as they need class labels or target values for training, but Clustering doesn't need target values.

QUESTION 2:

As the amount of training data increases

A. Training error usually decreases and generalization error usually increases
B. Training error usually decreases and generalization error usually decreases
C. Training error usually increases and generalization error usually decreases
D. Training error usually increases and generalization error usually increases

Correct Answer: A. Training error usually decreases and generalization error usually increases

Detailed Solution: When the training data increases, the decision boundary becomes very complex to fit the data. So, the generalization capability usually reduces with the increase in training data.

QUESTION 3:

Suppose I have 10,000 emails in my mailbox out of which 300 are spams. The spam detection system detects 150 mails as spams, out of which 50 are actually spams. What is the precision and recall of my spam detection system ?

A. **Precision = 33.33%, Recall = 16.66%**
B. Precision = 25%, Recall = 33.33%
C. Precision = 33.33%, Recall = 75%
D. Precision = 75%, Recall = 33.33%

Correct Answer: A. Precision = 33.33%, Recall = 16.66%

Detailed Solution:

Precision = $\frac{T_p}{T_p+F_p} = \frac{50}{50+100} = \frac{50}{150} = 33.33\%$

Recall = $\frac{T_p}{T_p+F_n} = \frac{50}{50+250} = 16.66\%$

QUESTION 4:

Which of the following are not classification tasks ?

A. Find the gender of a person by analyzing his writing style
B. **Predict the price of a house based on floor area, number of rooms etc**.
C. Predict whether there will be abnormally heavy rainfall next year
D. Detect Pneumonia from Chest X-ray images

Correct Answer: B. Predict the price of a house based on floor area, number of rooms etc.

Detailed Solution : House Price is a continuous real valued variable, so we have to use regression methods for this task.

QUESTION 5:

Occam's razor is an example of:

A. Inductive bias
B. Preference bias

Correct Answer: B. Preference bias

Detailed Solution : Prefer simplest hypothesis over complex one

---

QUESTION 6:

A feature F1 can take certain value: A, B, C, D, E, F and represents grade of students from a college. Which of the following statements is true in the following case?

A. Feature F1 is an example of a nominal variable.
B. Feature F1 is an example of ordinal variables.
C. It doesn't belong to any of the above categories.
D. Both of these

Correct Answer: B. Feature F1 is an example of ordinal variables.

Detailed Solution : Ordinal variables are the variables which have some order in their categories. For example, grade A should be considered as higher grade than grade B.

QUESTION 7:

Which of the following is a categorical feature?

A. Height of a person
B. Price of petroleum
C. Mother tongue of a person
D. Amount of rainfall in a day

Correct Answer: C. Mother tongue of a person

Detailed Solution : Categorical variables represent types of data which may be divided into groups. All other features are continuous)

QUESTION 8:

Which of the following tasks is NOT a suitable machine learning task?

A. **Finding the shortest path between a pair of nodes in a graph**
B. Predicting if a stock price will rise or fall
C. Predicting the price of petroleum
D. Grouping mails as spams or non-spams

Correct Answer : A. Finding the shortest path between a pair of nodes in a graph

Detailed Solution : Finding the shortest path is a graph theory based task, whereas other options are completely suitable for machine learning.

QUESTION 9:

Which of the following is correct for reinforcement learning?

A. The algorithm plans a sequence of actions from the current state.
B. The algorithm plans one action at each time step.
C. The training instances contain examples of states and best actions of the states.
D. The algorithm groups unseen data based on similarity.

Correct Answer : B.The algorithm plans one action at each time step.

Detailed Solution : In reinforcement learning, the agent tries to learn the policy function i.e. the best action to take at a given state.

QUESTION 10:

What is the use of Validation dataset in Machine Learning?

A. To train the machine learning model.
B. To evaluate the performance of the machine learning model
C. To tune the hyperparameters of the machine learning model
D. None of the above.

Correct Answer : C. To tune the hyperparameters of the machine learning model

Detailed Solution : The validation dataset provides an unbiased evaluation of a model fit on the training dataset while tuning the model's hyperparameters.

*********END*********