



Introduction to

Machine Learning

Assignment- Week 2

TYPE OF QUESTION: MCQ

Number of questions: 10

Total mark: 10 X 2 = 20

MCQ Question

QUESTION 1:

In a binary classification problem, out of 30 data points 12 belong to class I and 18 belong to class II. What is the entropy of the data set?

- A. 0.97
- B. 0
- C. 1
- D. 0.67

Answer: A. 0.97

Detailed Solution:

$$\text{Entropy} = - \left(\left(\frac{12}{30} \right) \log_2 \left(\frac{12}{30} \right) + \left(\frac{18}{30} \right) \log_2 \left(\frac{18}{30} \right) \right) = 0.97$$

QUESTION 2:

Which of the following properties are characteristics of decision trees?

- A. Low bias
- B. High variance
- C. Lack of smoothness of prediction surfaces
- D. None of the above

Correct Answer: A, B, C

Detailed Solution: Decision tree classifiers have low bias and high variance. As decision trees split the input space into rectangular spaces, the predictor surface or the decision boundary lacks smoothness.



QUESTION 3:

Statement: Decision Tree is an unsupervised learning algorithm.

Reason: The splitting criterion uses only the features of the data to calculate their respective measures.

- A. Statement is True. Reason is True.
- B. Statement is True. Reason is False.
- C. Statement is False. Reason is True.
- D. Statement is False. Reason is False.

Correct Answer: D. Statement is False. Reason is False.

Detailed Solution : Decision Tree is a supervised learning algorithm and the reason is also false.

QUESTION 4:

In linear regression, our hypothesis is $h_{\theta}(x) = \theta_0 + \theta_1 x$, the training data is given in the table.

x	y
10	5
3	3
6	7
8	6

If the cost function is $J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x_i) - y_i)^2$, where m is no. of training data points.

What is the value of $J(\theta)$ when $\theta = (1,1)$?

- A. 0
- B. 5.75
- C. 4.75
- D. 6.75

Correct Answer: B. 5.75

Detailed Solution : Substitute θ_0 by 1 and θ_1 by 1 and compute $J(\theta)$.



QUESTION 5:

What is a common indicator of overfitting in a decision tree?

- A. The training accuracy is high while the test accuracy is low.
- B. The tree is shallow.
- C. The tree has only a few leaf nodes.
- D. The tree's depth matches the number of attributes in the dataset.

Correct Answer: A. The training accuracy is high while the test accuracy is low.

Detailed Solution: The training accuracy is high while the test accuracy is low.

QUESTION 6:

What is true for Batch Gradient Descent?

- A. In every iteration, model parameters are updated based on one training sample
- B. In every iteration, model parameters are updated based on all training samples
- C. None of the above



Correct Answer : B. In every iteration model parameters are updated based on all training samples.

Detailed Solution : In batch gradient descent, all training samples are used in every iteration. In stochastic gradient descent, one training sample is used to update parameters in every iteration.

QUESTION 7:

Answer Questions 7-8 with the data given below:

Consider the following dataset. We want to build a decision tree classifier to detect whether a tumor is malignant or not using several input features such as age, vaccination, tumor size and tumor site. The target variable is “Malignant” and the other attributes are input features.

Age	Vaccination	Tumor Size	Tumor Site	Malignant
5	1	Small	Shoulder	0
9	1	Small	Knee	0
6	0	Small	Marrow	0
6	1	Medium	Chest	0
7	0	Medium	Shoulder	0
8	1	Large	Shoulder	0
5	1	Large	Liver	0
9	0	Small	Liver	1
8	0	Medium	Shoulder	1
8	0	Medium	Shoulder	1
6	0	Small	Marrow	1
7	0	Small	Chest	1

What is the initial entropy of the dataset?

- A. 0.543
- B. 0.9798
- C. 0.8732
- D. 1



Correct Answer: B. 0.9798

Detailed Solution:

The entropy of the whole dataset is $= -(5/12)\log_2(5/12) - (7/12)\log_2(7/12) = 0.9798$

QUESTION 8:

For the dataset in Question 7, what is the information gain of Vaccination (If entropy measure is used to calculate information gain)?

- A. 0.4763
- B. 0.2102
- C. 0.1134
- D. 0.9355

Correct Answer: A. 0.4763

Information gain of Vaccination =

$$0.9798 - \left(\frac{5}{12} \left(-\frac{0}{5} \log_2 \frac{0}{5} - \frac{5}{5} \log_2 \frac{5}{5} \right) - \frac{7}{12} \left(-\frac{2}{7} \log_2 \frac{2}{7} - \frac{5}{7} \log_2 \frac{5}{7} \right) \right) = 0.4763$$

QUESTION 9:

Which of the following criteria is typically used for optimizing in linear regression?

- A. Maximizing the number of points touched by the line
- B. Minimizing the number of points touched by the line
- C. Minimizing the sum of squared distance of the line from the points
- D. Minimizing the maximum squared distance of a point from a line

Correct Answer: C. Minimizing the sum of squared distance of the line from the points

Detailed Solution: In linear regression, the objective is to minimize the sum of squared distance of the line from the points.

QUESTION 10:

The parameters obtained in linear regression

- A. can take any value in the real space
- B. are strictly integers
- C. always lie in the range $[0,1]$
- D. can take only non-zero values

Correct Answer: A. can take any value in the real space

Detailed Solution: The linear regression parameters can take any real number value.

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