NPTEL Online Certification Courses Indian Institute of Technology Kharagpur

Introduction to
Machine Learning
Assignment- Week 2
TYPE OF QUESTION: MCQ

Number of questions: **8** **Total mark: 8 X 2 = 16**

MCQ Question

QUESTION 1:

Identify whether the following statement is true or false?

“Overfitting is more likely when the set of training data is small”

A. **True**
B. False

Correct Answer : **A.True**

Detailed Solution : With a small training dataset, it’s easier to find a hypothesis to fit the training data exactly,i.e., overfit.

QUESTION 2:

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

A. Maximize the number of points it touches.
B. Minimize the number of points it touches.
C. **Minimize the squared distance from the points.**
D. Minimize the maximum distance of a point from a line.

Correct Answer : **C. Minimize the squared distance from the points.**

Detailed Solution : Loss function of linear regression is squared distance from the points.

QUESTION 3:

Which of the following is false?

A. Bias is the true error of the best classifier in the concept class
B. Bias is high if the concept class cannot model the true data distribution well
C. **High bias leads to overfitting**
D. For high bias both train and test error will be high

Correct Answer : C. High bias leads to overfitting

Detailed Solution : High bias leads to underfitting.

QUESTION 4:

The following dataset will be used to learn a decision tree for predicting whether a person is happy (H) or sad (S), based on the color of shoes, whether they wear a wig and the number of ears they have.

Color	Wig	Num. Ears	Emotion (Output)
G	Y	2	S
G	N	2	S
G	N	2	S
B	N	2	S
B	N	2	H
R	N	2	H
R	N	2	H
R	N	2	H
R	Y	3	H

Which attribute should you choose as the root of the decision tree?

A. Color

B. Wig

C. Number of ears

D. Any one of the previous three attributes

Correct Answer : A. Color

Detailed Solution : We have to compute Information Gain w.r.t. each of these 4 attributes and the attribute with highest information gain will be chosen as the root of the decision tree.

QUESTION 5:

Consider applying linear regression with the hypothesis as $h_\theta(x) = \theta_0 + \theta_1 x$.The training data is given in the table.

X	Y
6	7
5	4
10	9
3	4

The cost function is $J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x_i) - y_i)^2$

What is the value of $J(\theta)$when **θ** = (2, 1) ?

A. 0

B. 1

C. 2

D. 2.5

Correct Answer: D. 2.5

Detailed Solution : Substitute θ_0 by 2 and θ_1 by 1 and compute $J(\theta)$.

QUESTION 6:

In a binary classification problem, out of 64 data points 29 belong to class I and 35 belong to class II. What is the entropy of the data set?

A. 0.97
B. 0
C. 1
D. 0.99

Correct Answer : D. 0.99

Detailed Solution : We can compute Entropy as

$ENTROPY(p_+, p_-) = -p_+ log_2 p_+ - p_- log_2 p_-$, here

$p_+ = 29/64$ *and* $p_- = 35/64$

QUESTION 7:

Decision trees can be used for the following type of datasets:

I. The attributes are categorical
II. The attributes are numeric valued and continuous
III. The attributes are discrete valued numbers

A. In case I only
B. In case II only
C. In cases II and III only
D. In cases I, II and III

Correct Answer : D. In cases I, II and III

Detailed Solution : Decision trees can be applied in all 3 cases.

QUESTION 8:

What is true for Stochastic Gradient Descent?

A. In every iteration, model parameters are updated for multiple training samples
B. In every iteration, model parameters are updated for one training sample
C. In every iteration, model parameters are updated for all training samples
D. None of the above

Correct Answer : B. In every iteration model parameters are updated for one training sample.

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

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