NPTEL Online Certification Courses

Indian Institute of Technology Kharagpur

Introduction to Machine Learning

Assignment- Week 3

TYPE OF QUESTION: MCQ

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

QUESTION 1:

Suppose, you have been given the following data where x1 and x2 are the 2 input variables and Class is the dependent variable.

x1	x2	Class
-1	1	-
0	1	+
0	2	-
1	-1	-
1	0	+
1	2	+
2	2	-
2	3	+

What will be the class of a new data point x1=1 and x2=1 in 5-NN (k nearest neighbour with k=5) using euclidean distance measure?

A. + Class
B. – Class
C. Cannot be determined

Correct Answer: **A. + Class**
Detailed Solution : 5 nearest points to the new point (1,1) are: (0,1), (0,2), (1,0), (1,2), (2,2). The majority class among these 5 nearest neighbours is + Class.

QUESTION 2:

Imagine you are dealing with a 10 class classification problem. What is the maximum number of discriminant vectors that can be produced by LDA?

A. 20
B. 14
C. 9
D. 10

Correct Answer: C. 9

Detailed Solution : LDA produces at most c − 1 discriminant vectors, c = no of classes

---

QUESTION 3:

Fill in the blanks:

K-Nearest Neighbor is a _____ , _____ algorithm

A. Non-parametric, eager
B. Parametric, eager
C. Non-parametric, lazy
D. Parametric, lazy

Correct Answer: C. Non-parametric, lazy

Detailed Solution: KNN is non-parametric because it does not make any assumption regarding the underlying data distribution. It is a lazy learning technique because during training time it just memorizes the data and finally computes the distance during testing.

QUESTION 4:

Which of the following statements is True about the KNN algorithm?

A. KNN algorithm does more computation on test time rather than train time.
B. KNN algorithm does lesser computation on test time rather than train time.
C. KNN algorithm does an equal amount of computation on test time and train time.
D. None of these.

Correct Answer: A. KNN algorithm does more computation on test time rather than train time.

Detailed Solution : The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.
In the testing phase, a test point is classified by assigning the label which is the most frequent among the k training samples nearest to that query point – hence higher computation.

QUESTION 5:

Which of the following necessitates feature reduction in machine learning?

1. Irrelevant and redundant features
2. Curse of dimensionality
3. Limited computational resources.

A. 1 only
B. 2 only
C. 1 and 2 only
D. 1, 2 and 3

Correct Answer: D. 1,2 and 3

Detailed Solution: All these things necessitate feature reduction.

QUESTION 6:

When there is noise in data, which of the following options would improve the performance of the k-NN algorithm?

A. Increase the value of k
B. Decrease the value of k
C. Changing value of k will not change the effect of the noise
D. None of these

Correct Answer: A. Increase the value of k

Detailed Solution : Increasing the value of k reduces the effect of the noise and improves the performance of the algorithm.

QUESTION 7:

Find the value of the Pearson's correlation coefficient of X and Y from the data in the following table.

AGE (X)	GLUCOSE (Y)
43	99
21	65
25	79
42	75

A. 0.47
B. 0.68
C. 1
D. 0.33

Correct Answer : B. 0.68

Detailed Solution : Pearson Coefficient $r = \frac{\sum_i (X_i - \bar{X})((Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2}\sqrt{\sum_i (Y_i - \bar{Y})^2}}$

Where X = [43,21,25,42], Y = [99,65,79,75], $\overline{X}$ = mean of X_i values and $\overline{Y}$ = mean of Y_i values.

QUESTION 8:

Which of the following statements is/are true about PCA?

1. PCA is a supervised method
2. It identifies the directions that data have the largest variance
3. Maximum number of principal components <= number of features
4. All principal components are orthogonal to each other

A. Only 2
B. 1, 3 and 4
C. 1, 2 and 3
D. 2, 3 and 4

Correct Answer: D

Detailed Solution : PCA is an unsupervised learning algorithm, so 1 is wrong. Other statements are true about PCA.

QUESTION 9:

In user-based collaborative filtering based recommendation, the items are recommended based on :

A. Similar users
B. Similar items
C. Both of the above
D. None of the above

Correct Answer: A. Similar users

Detailed Solution: In User-based Collaborative filtering, items are recommended based on similar users.

QUESTION 10:

Identify whether the following statement is true or false?
"*Linear Discriminant Analysis (LDA) is a supervised method*"

A. TRUE
B. FALSE

Correct Answer : A. TRUE

Detailed Solution : LDA is a supervised method as it makes use of the class labels.

******END****