NPTEL Online Certification Courses Indian Institute of Technology Kharagpur

Introduction to Machine Learning

Assignment- Week 3

TYPE OF QUESTION: MCQ

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

QUESTION 1:

Suppose, you have given the following data where x and y are the 2 input variables and Class is the dependent variable.

X	Y	Class
-1	1	-
0	1	+
0	2	-
1	-1	-
1	0	+
1	2	+
2	2	-
2	3	+

Suppose, you want to predict the class of new data point x=1 and y=1 using euclidean distance in 7-NN. To which class the data point belongs to?

A. + Class
B. – Class
C. Can't say
D. None of these

Correct Answer: **B. – Class**

Detailed Solution : We have to compute the euclidean distance from the given point (1,1) to all the data points given in the dataset and based on that we have to check the dominating class for the 7 nearest points.

QUESTION 2:

Imagine you are dealing with 15 class classification problem. What is the maximum number of discriminant vectors that can be produced by LDA?

A. 20
B. 14
C. 21
D. 10

Correct Answer: B. 14

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

QUESTION 3:

'People who bought this, also bought…' recommendations seen on amazon is a result of which algorithm?

A. User based Collaborative filtering
B. Content based filtering
C. Item based Collaborative filtering
D. None of the above

Correct Answer: C. Item based Collaborative filtering

Detailed Solution : Though both User based and Item based CF methods are used in recommendation systems, Amazon specifically uses Item based filtering.

QUESTION 4:

Which of the following 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

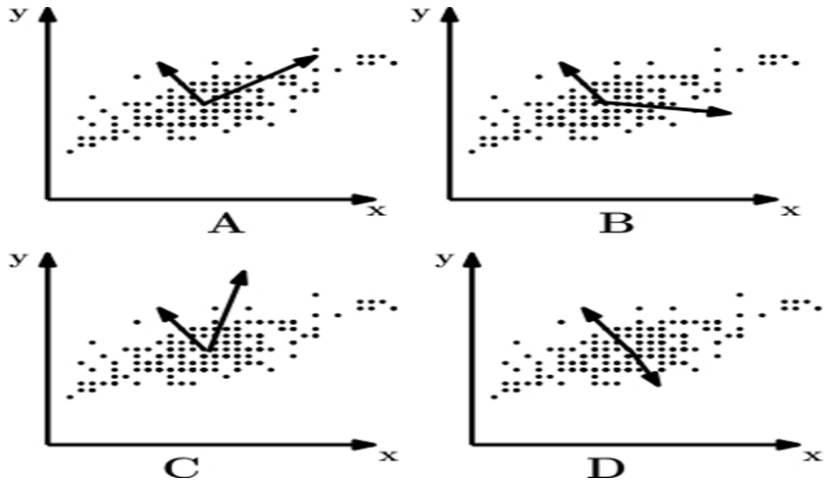
A. Only 2
B. 1, 3 and 4
C. 1, 2 and 3
D. 2, 3 and 4

Correct Answer: D

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

QUESTION 5:

Consider the figures below. Which figure shows the most probable PCA component directions for the data points?



A. A

B. B

C. C

D. D

Correct Answer: **A. A**

Detailed Solution : PCA tries to choose the direction in such a way that maximizes the variance in the data.

QUESTION 6:

When there is noise in data, which of the following options would improve the performance of the KNN 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:

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 are most frequent among the k training samples nearest to that query point – hence higher computation.

QUESTION 8:

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 $r = \frac{\sum_i (X_i - \overline{X})((Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2}\sqrt{\sum_i (Y_i - \overline{Y})^2}}$

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