Introduction to Machine Learning -IITKGP

Assignment - 3

TYPE OF QUESTION: MCQ/MSQ

Number of questions: 15 Total mark: 2 * 15 = 30

Q1. 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

Explanation: 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.

Q2. You have been given the following 2 statements. Find out which of these options is/are true in the case of k-NN.

(i) In case of very large value of k, we may include points from other classes into the neighborhood.

(ii) In case of too small value of k, the algorithm is very sensitive to noise.

a. (i) is True and (ii) is False
b. (i) is False and (ii) is True
c. Both are True
d. Both are False

Correct Answer: c

Explanation: Both options are true and are self-explanatory.

Q3. State whether the statement is True/False:

k-NN algorithm does more computation on test time rather than train time.

a. True
b. False

Correct Answer: a

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

Q4. Suppose you are given the following images (1 represents the left image, 2 represents the middle and 3 represents the right). Now your task is to find out the value of k in k-NN in each of the images shown below. Here k1 is for 1st, k2 is for 2nd and k3 is for 3rd figure.

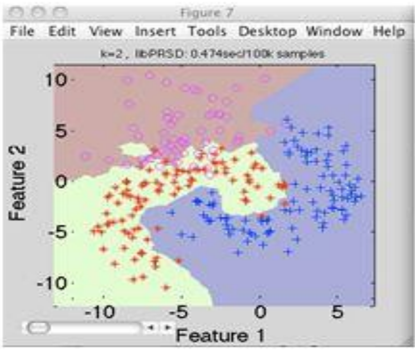


Figure 7
File Edit View Insert Tools Desktop Window Help
k=2, libPRSD: 0.474sec/100k samples
10
5
0
-5
-10
Feature 2
-10
-5
0
5
Feature 1

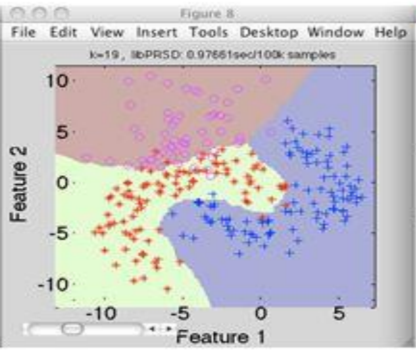


Figure 8
File Edit View Insert Tools Desktop Window Help
k=19, libPRSD: 0.97661sec/100k samples
10
5
0
-5
-10
Feature 2
-10
-5
0
5
Feature 1

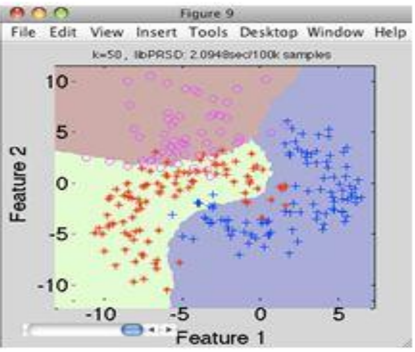


Figure 9
File Edit View Insert Tools Desktop Window Help
k=50, libPRSD: 2.0948sec/100k samples
10
5
0
-5
-10
Feature 2
-10
-5
0
5
Feature 1

a. k1 > k2> k3
b. k1 < k2> k3
c. k1 < k2 < k3
d. None of these

Correct Answer: c

Explanation: The value of k is highest in k3, whereas in k1 it is lowest. As the decision boundary is more smooth in the right image than the others.

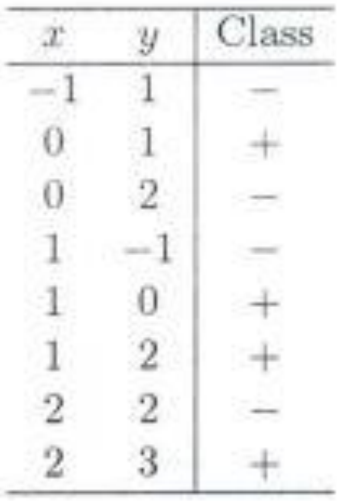
Q5. Which of the following necessitates feature reduction in machine learning?

a. Irrelevant and redundant features
b. Limited training data
c. Limited computational resources.
d. All of the above

Correct Answer: d

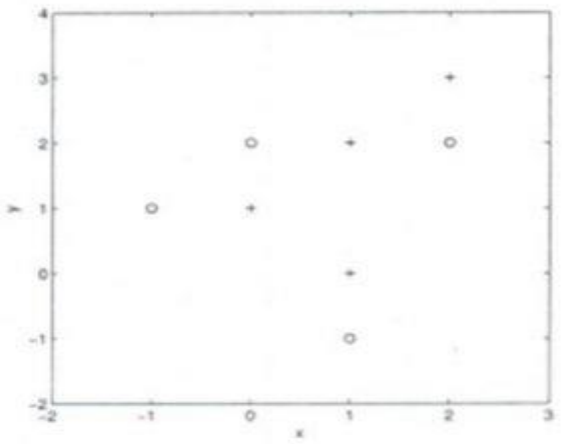
Detailed Solution: Follow lecture slides

Q6. 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	$+$

Below is a scatter plot which shows the above data in 2D space.



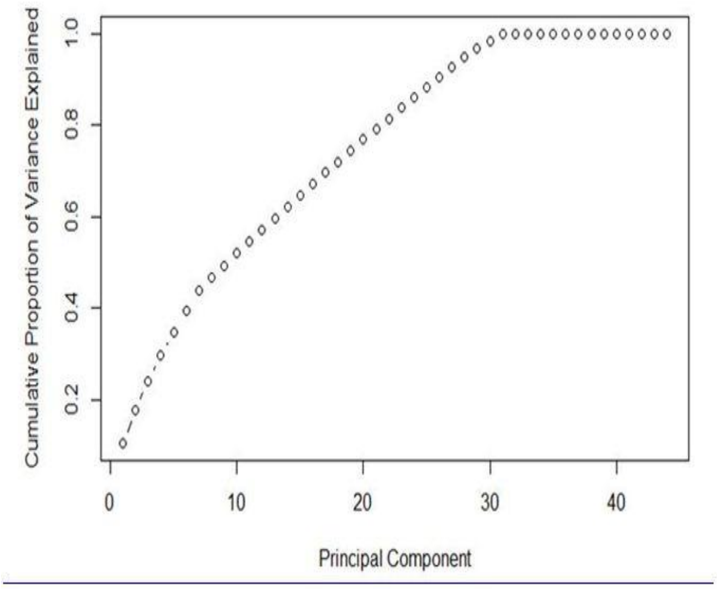
Suppose, you want to predict the class of new data point x=1 and y=1 using Euclidean distance in 3-NN. In which class this data point belongs to?

a. + Class
b. – Class
c. Can't Say
d. None of these

Correct Answer: a

Explanation: All three nearest point are of + class so this point will be classified as + class.

Q7. What is the optimum number of principal components in the below figure?



a. 10
b. 20
c. 30
d. 40

Correct Answer: c

Explanation: We can see in the above figure that the number of components = 30 is giving highest variance with lowest number of components. Hence option 'c' is the right answer.

Q8. Suppose we are using dimensionality reduction as pre-processing technique, i.e, instead of using all the features, we reduce the data to k dimensions with PCA. And then use these PCA projections as our features. Which of the following statements is correct?

Choose which of the options is correct?

a. Higher value of ‘k’ means more regularization
b. Higher value of ‘k’ means less regularization

Correct Answer: b

Explanation: The higher value of ‘k’ would lead to less smoothening of the decision boundary. This would be able to preserve more characteristics in data, hence less regularization.

Q9. In 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

Explanation: Follow the definition of collaborative filtering.

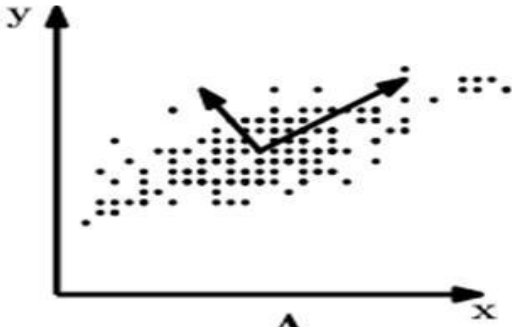
Q10. The major limitation of collaborative filtering is:

a. Cold start
b. Overspecialization
c. None of the above

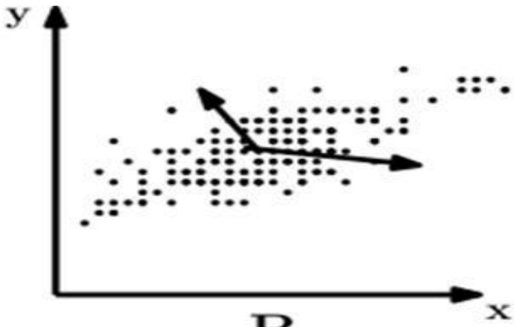
Correct Answer: a

Explanation: For new users, we have very few transactions. So, it’s very difficult to find similar users.

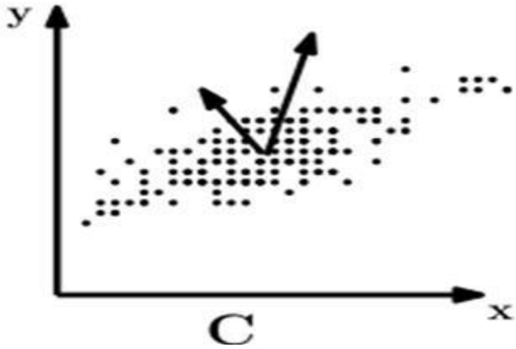
Q11. Consider the figures below. Which figure shows the most probable PCA component directions for the data points?



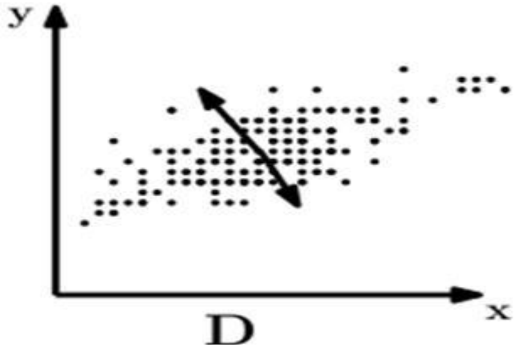
A



B



C



D

a. A
b. B
c. C
d. D

Correct Answer: a

Explanation: [Follow the lecture slides]

Choose directions such that a total variance of data will be maximum

1. Maximize Total Variance

Choose directions that are orthogonal

2. Minimize correlation

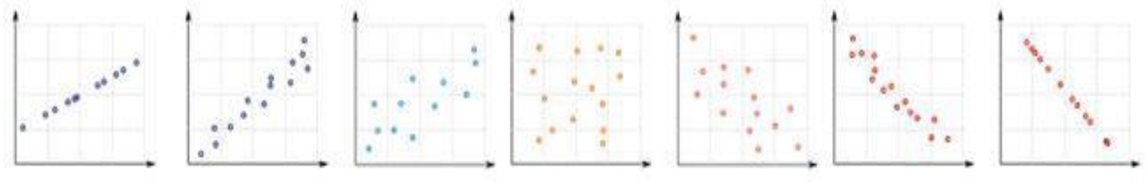
Q12. Suppose that you wish to reduce the number of dimensions of a given data to dimensions using PCA. Which of the following statement is correct?

a. Higher means more regularization
b. Higher means less regularization
c. Can’t Say

Correct Answer: b

Explanation: Higher k value leads to less smoothening as we preserve more characteristics in data, hence less regularization.

Q13. Suppose you are given 7 plots 1-7 (left to right) and you want to compare Pearson correlation coefficients between variables of each plot. Which of the following is true?



1. 1<2<3<4
2. 1>2>3>4
3. 7<6<5<4
4. 7>6>5>4

a. 1 and 3
b. 2 and 3
c. 1 and 4
d. 2 and 4

Correct Answer: b

Explanation: From image 1 to 4, correlation is decreasing (coefficient values are positive). From image 4 to 7 correlation is increasing, but the coefficient values are negative (for example, 0, -0.3, -0.7, -0.99).

Q14. Imagine you are dealing with 20 class classification problem. What is the maximum number of discriminant vectors that can be produced by LDA?
a. 20
b. 19
c. 21
d. 10

Correct Answer: b
Explanation: LDA produces at most c − 1 discriminant vectors.

Q15. In which of the following situations collaborative filtering algorithm is appropriate?
a. You manage an online bookstore and you have the book ratings from many users. For each user, you want to recommend other books he/she will like based on her previous ratings and other users' ratings.

b. You manage an online bookstore and you have the book ratings from many users. You want to predict the expected sales volume (No of books sold) as a function of average rating of a book.

c. Both A and B

d. None of the above

Correct Answer: a

Explanation: Collaborative filtering is a recommendation technique that is specifically designed for situations like the one described in option a. In collaborative filtering, recommendations are made based on the patterns of user preferences and behaviors. It analyzes the historical data of user-item interactions, such as book ratings given by users, to find similarities between users and items.

Option b is not appropriate for collaborative filtering because it involves predicting the expected sales volume (number of books sold) as a function of average rating of a book. Collaborative filtering focuses on user-item interactions and is more concerned with providing personalized recommendations rather than predicting sales volume based on average ratings.

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