Assignment – 3

1. In k-NN algorithm, given a set of training examples and the value of k < size of training set (n), the algorithm predicts the class of a test example to be the
   A) **Most frequent class among the classes of k closest training examples**.
   B) Least frequent class among the classes of k closest training examples.
   C) Class of the closest point.
   D) Most frequent class among the classes of the k farthest training examples.

   Solution: A (From the k-means algorithm. Follow the lecture slides)

2. In collaborative Filtering based Recommendation, the items are recommended based on which of the following?
   A) Similar users
   B) Similar items
   C) Both A and B
   D) None

   Solution: A (From the definition of Collaborative Filtering)

3. Which of the following are advantages of large value of k in k-NN algorithm?
   A) Less sensitive to noise.
   B) Better probability estimates for discrete classes.
   C) Larger training sets allow larger values of k.
   D) All of the above.

   Solution: D (From lecture slides)

4. 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.

Solution: D (From lecture slides)

5. For which of the following cases Dimensional reduction may be used?
   A) Data Compression
   B) Data Visualization
   C) To prevent overfitting
   D) Both A and B

   Solution: D (From lecture slides)

6. Which of the following is the limitation of Collaborative Filtering?
   A) Over specialization
   B) Cold start
   C) Both A and B
   D) None

   Solution: B (For new users, we have very few transactions. So, it's difficult to find similar users.)

7. **Which of the following statements is true about PCA?**

   (i) We must standardize the data before applying PCA.

   (ii) We should select the principal components which explain the highest variance

   (iii) We should select the principal components which explain the lowest variance

   (iv) We can use PCA for visualizing the data in lower dimensions

   A. (i), (ii) and (iv)

   B. (ii) and (iv)

   C. (iii) and (iv)

   D. (i) and (iii)

   Solution: A (From lecture slides)

8. In feature selection, which of the following techniques can be used to find a subset of features?

   A) Sequential forward search

   B) Sequential backward search

   C) Both A and B

   D) None of A or B

   Solution: C (A and B are feature selection approaches)

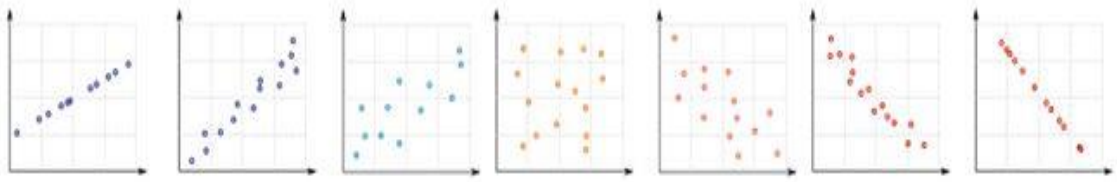
9. [True or False] A Pearson correlation between two variables is zero but, still their values can still be related to each other.

   A) TRUE

   B) FALSE

   Solution: (A) (*Y*=*X*2. Note that, they are not only associated, but one is a function of the other and Pearson correlation between them is 0.)

10. Suppose you are given 7 Scatter plots 1-7 (left to right) and you want to compare Pearson correlation coefficients between variables of each scatterplot. Which of the following is in the right order?



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

Solution: (B) (from image 1to 4 correlation is decreasing (absolute value). But, from image 4 to 7 correlation is increasing but values are negative (for example, 0, -0.3, -0.7, -0.99).)