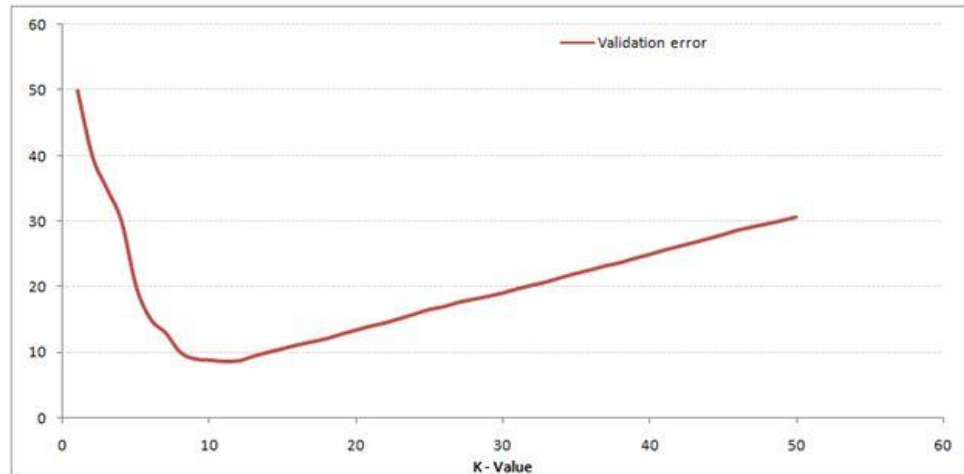1. k-NN algorithm does more computation on test time rather than train time.

    A) TRUE
    B) FALSE

2. In the image below, which would be the best value for k assuming that the algorithm you are using is k-Nearest Neighbor.

    A) 3
    B) 10
    C) 20
    D 50

3. Which of the following distance metric can not be used in k-NN?
        A) Manhattan
        B) Minkowski
        C) Tanimoto
        D) Jaccard
        E) Mahalanobis
        F) All can be used

4. Which of the following option is true about k-NN algorithm?
        A) It can be used for classification
        B) It can be used for regression
        C) It can be used in both classification and regression

5. Which of the following statement is true about k-NN algorithm?
        1. k-NN performs much better if all of the data have the same scale
        2. k-NN works well with a small number of input variables (p), but struggles when the number of inputs is very large
        3. k-NN makes no assumptions about the functional form of the problem being solved

A) 1 and 2
B) 1 and 3
C) Only 1
D) All of the above

6. Which of the following machine learning algorithm can be used for imputing missing values of both categorical and continuous variables?
A) K-NN
B) Linear Regression
C) Logistic Regression

7. Which of the following is true about Manhattan distance?
A) It can be used for continuous variables
B) It can be used for categorical variables
C) It can be used for categorical as well as continuous
D) None of these

8. Which of the following distance measure do we use in case of categorical variables in k-NN?
1. Hamming Distance
2. Euclidean Distance
3. Manhattan Distance

A) 1
B) 2
C) 3
D) 1 and 2
E) 2 and 3
F) 1,2 and 3

9. Which of the following will be Euclidean Distance between the two data point A(1,3) and B(2,3)?
A) 1
B) 2
C) 4
D) 8

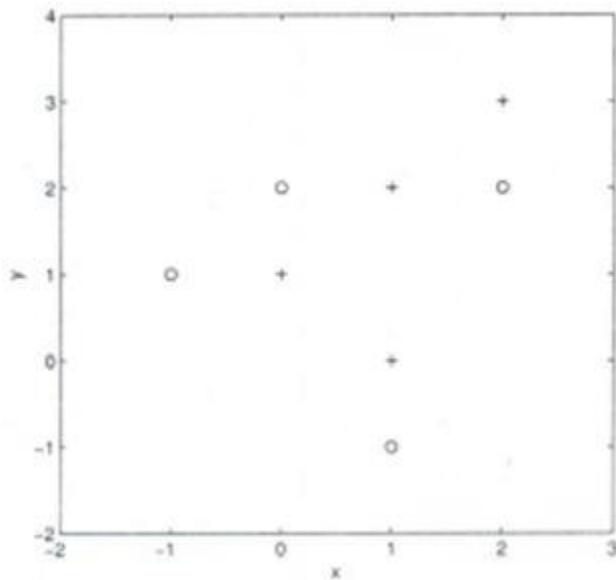10. Which of the following will be Manhattan Distance between the two data point A(1,3) and B(2,3)?
A) 1
B) 2
C) 4
D) 8

Context: 11-12

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.

11. Suppose, you want to predict the class of new data point x=1 and y=1 using eucludian distance in 3-NN. In which class this data point belong to?
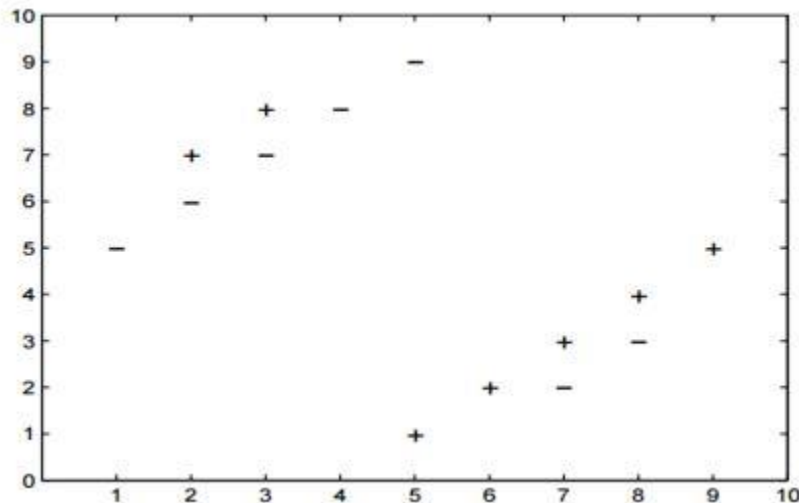
    A) + Class
    B) – Class
    C) Can't say
    D) None of these

12. In the previous question, you are now want use 7-NN instead of 3-KNN which of the following x=1 and y=1 will belong to?

    A) + Class
    B) – Class
    C) Can't say

Context 13-14:
Suppose you have given the following 2-class data where "+" represent a postive class and "" is represent negative class.

13. Which of the following value of k in k-NN would minimize the leave one out cross validation accuracy?
    A) 3
    B) 5
    C) Both have same
    D) None of these

14. Which of the following would be the leave on out cross validation accuracy for k=5?
    A) 2/14
    B) 4/14
    C) 6/14
    D) 8/14
    E) None of the above

15. Which of the following will be true about k in k-NN in terms of Bias?
    A) When you increase the k the bias will be increases
    B) When you decrease the k the bias will be increases
    C) Can't say
    D) None of these

16. Which of the following will be true about k in k-NN in terms of variance?
    A) When you increase the k the variance will increases
    B) When you decrease the k the variance will increases
    C) Can't say
    D) None of these

17. The following two distances(Eucludean Distance and Manhattan Distance) have given to you which generally we used in K-NN algorithm. These distance are between two points A(x1,y1) and B(x2,Y2).

Your task is to tag the both distance by seeing the following two graphs. Which of the following option is true about below graph ?

  A) Left is Manhattan Distance and right is euclidean Distance
  B) Left is Euclidean Distance and right is Manhattan Distance
  C) Neither left or right are a Manhattan Distance
  D) Neither left or right are a Euclidian Distance

18. When you find noise in data which of the following option would you consider in k-NN?
  A) I will increase the value of k
  B) I will decrease the value of k
  C) Noise can not be dependent on value of k
  D) None of these

19. In k-NN it is very likely to overfit due to the curse of dimensionality. Which of the following option would you consider to handle such problem?
 1. Dimensionality Reduction
  2. Feature selection

  A) 1
  B) 2
  C) 1 and 2
  D) None of these

20. Below are two statements given. Which of the following will be true both statements?
  1. k-NN is a memory-based approach is that the classifier immediately adapts as we collect new training data.
  2. The computational complexity for classifying new samples grows linearly with the number of samples in the training dataset in the worst-case scenario.

  A) 1
  B) 2
  C) 1 and 2
  D) None of these

21. Suppose you have given the following images(1 left, 2 middle and 3 right), Now your task is to find out the value of k in k-NN in each image where k1 is for 1st, k2 is for 2nd and k3 is for 3rd figure.
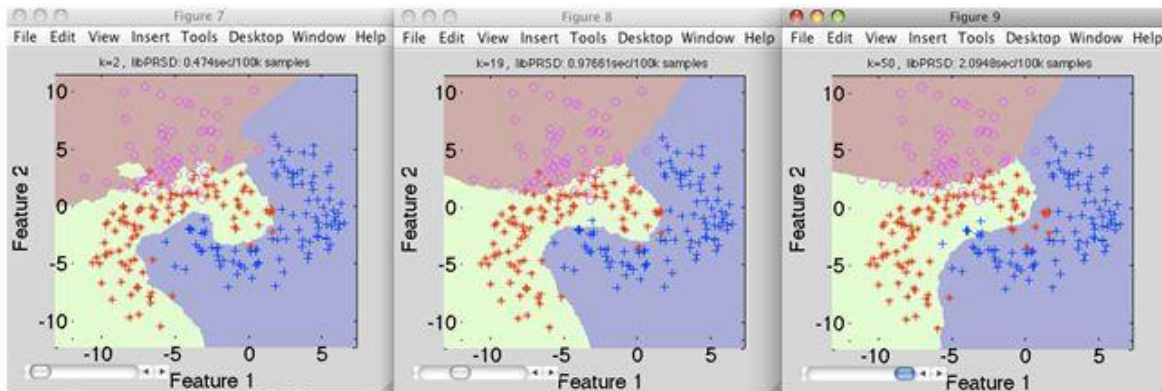
A) k1 > k2> k3
B) k1<k2
C) k1 = k2 = k3
D) None of these

22. Which of the following value of k in the following graph would you give least leave one out cross validation accuracy?

A) 1
B) 2
C) 3
D) 5

23. A company has build a kNN classifier that gets 100% accuracy on training data. When they deployed this model on client side it has been found that the model is not at all accurate. Which of the following thing might gone wrong?
Note: Model has successfully deployed and no technical issues are found at client side except the model performance

A) It is probably a overfitted model
B) It is probably a underfitted model

C) Can't say
D) None of these

24. You have given the following 2 statements, find which of these option is/are true in case of k-NN?

1. In case of very large value of k, we may include points from other classes into the neighborhood.
2. In case of too small value of k the algorithm is very sensitive to noise

A) 1
B) 2
C) 1 and 2
D) None of these

25. Which of the following statements is true for k-NN classifiers?

A) The classification accuracy is better with larger values of k
B) The decision boundary is smoother with smaller values of k
C) The decision boundary is linear
D) k-NN does not require an explicit training step

26. True-False: It is possible to construct a 2-NN classifier by using the 1-NN classifier?

A) TRUE
B) FALSE

27. In k-NN what will happen when you increase/decrease the value of k?

A) The boundary becomes smoother with increasing value of K
B) The boundary becomes smoother with decreasing value of K
C) Smoothness of boundary doesn't dependent on value of K
D) None of these

28. Following are the two statements given for k-NN algor thm, which of the statement(s) is/are true?

1. We can choose optimal value of k with the help of cross validation
2. Euclidean distance treats each feature as equally important

A) 1
B) 2
C) 1 and 2
D) None of these

Context 29-30:

Suppose, you have trained a k-NN model and now you want to get the prediction on test data. Before getting the prediction suppose you want to calculate the time taken by k-NN for predicting the class for test data.

Note: Calculating the distance between 2 observation will take D time.

29. What would be the time taken by 1-NN if there are N(Very large) observations in test data?

A) N*D
B) N*D*2
C) (N*D)/2
D) None of these

30. What would be the relation between the time taken by 1-NN,2-NN,3-NN.
A) 1-NN >2-NN >3-NN
B) 1-NN < 2-NN < 3-NN
C) 1-NN ~ 2-NN ~ 3-NN
D) None of these

https://www.analyticsvidhya.com/blog/2017/09/30-questions-test-k-nearest-neighbors-algorithm/