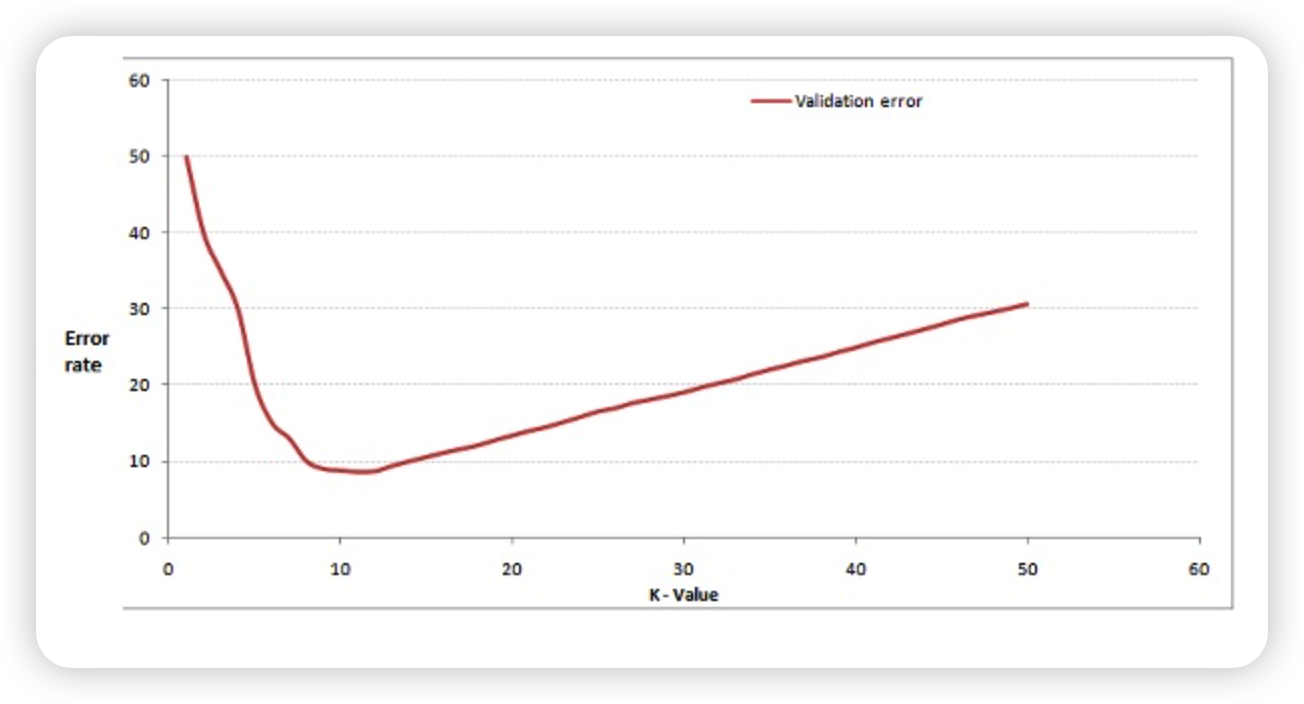
Assignment 3

Please write down you answers in the table. This makes it easier for me to mark. Thank you! Total mark for this assignment is 18. Each question is worth 2 points.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Questions | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Answer | B | C | D | A | B | A | A | B | Both |

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



1. 3
2. 10
3. 20
4. 50
5. **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.**

1. **Which of the following statement is true about k-NN algorithm?**
2. k-NN performs much better if all of the data have the same scale
3. k-NN works well with a small number of input variables, but struggles when the number of inputs is very large.
4. 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

**4.** 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

5.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

6.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

**7.** 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

Note: We did not cover Hamming Distance and Manhattan Distance in class. But you need to do some research yourself to figure out what is hamming distance and Manhattan distance to figure out this problem.

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

**8.** 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

9.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