UE Machine Learning: Supervised Techniques

Assignment 1 Report

Name: Omar Amr

Matrikel-Nr: k11776960

1) For the original data set, the best k was equal to 1 as shown below in Figure 1.

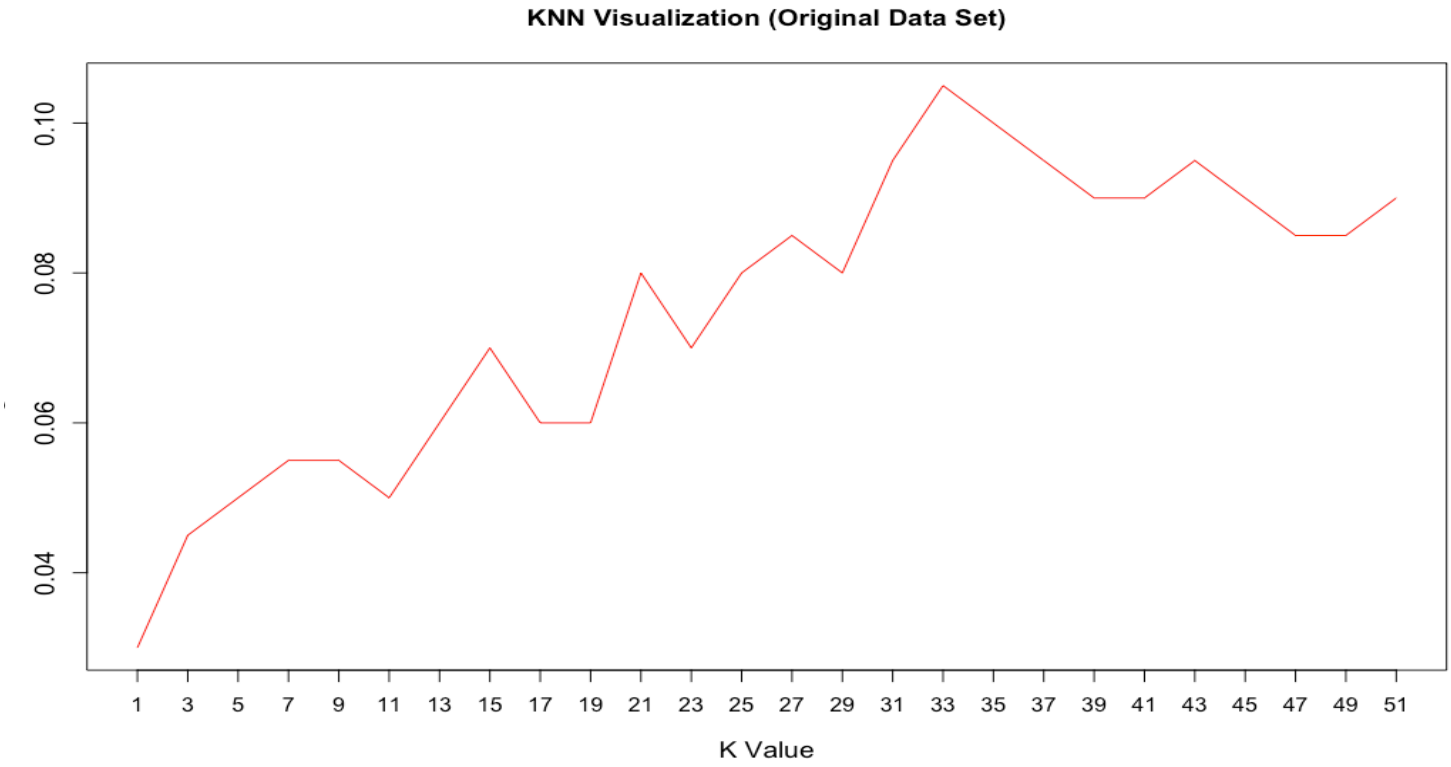


Figure 1: K vs Average Error Plotting for Original Data Set

Explanation:
As shown in the Figure 2 below, which represents the data visualization using the two features as indexes, the two classes can be separated into only 2 areas.

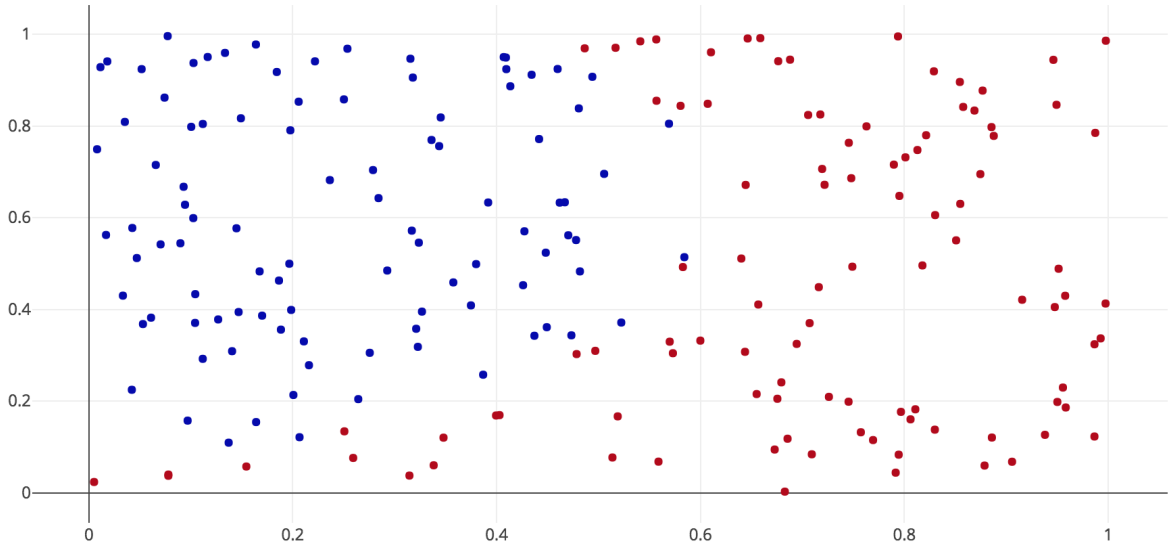


Figure 2: Original Data Set Visualization

For k=1, this means that any testing point will be classified according to only its very closest neighbor. Since the figure above shows that points of the same class are very close to each other and a line can be drawn to create two areas one for each class, this means judging the class of a testing point according to only its one very nearest neighbor is effective. When K is increased, a point will be judged according to the k nearest neighbors to it. This means a red point that is very close to another red point but also close enough to another two blue points will be classified as blue which is not correct.

2) After flipping the labels in the original data set with a probability of 20, the best k that has the least amount of error was equal to 45 as shown in Figure 2 below.

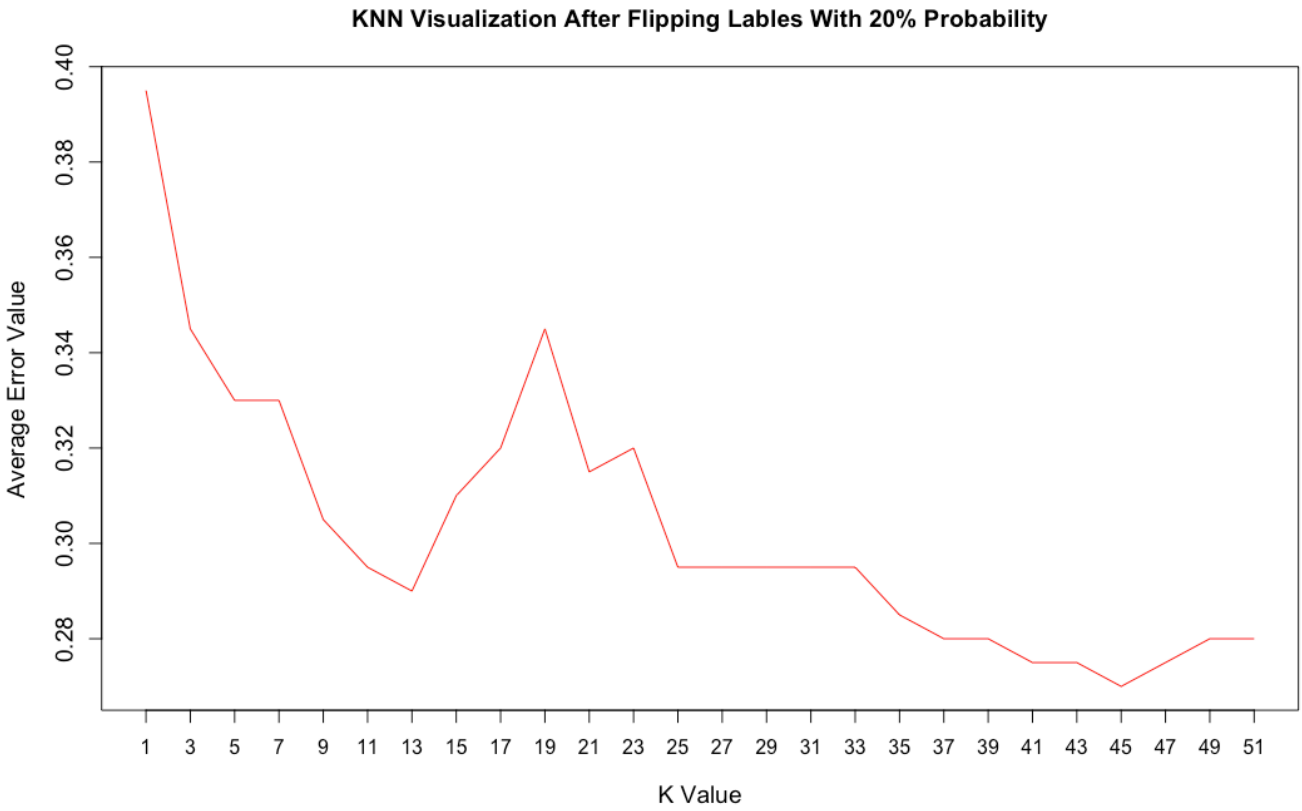


Figure 3: K vs Average Error After Flipping labels with 20% Probability

Explanation:
Figure 4 below shows the visualization of data after flipping labels with 20% probability.

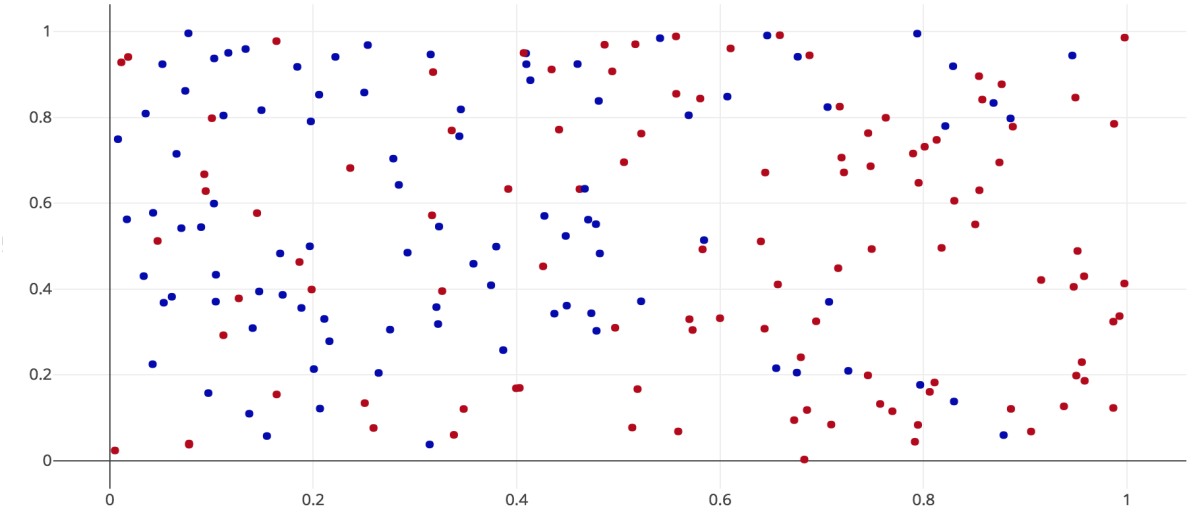
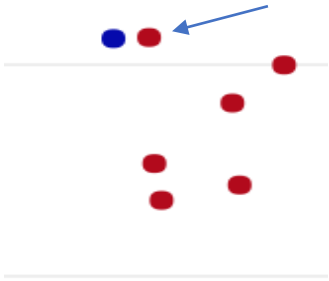


Figure 4: Flipped Data Set Visualization

Figure 4 shows that some of the blue points are now classified as red and vice-versa. This means classifying a point according to only its one nearest neighbor is not a good idea. For example, the red point highlighted with the arrow would be classified as blue if K = 1 is used. This explains the high error for low values of K.



With K=45, a point will be classified according to the majority of its 45 neighbors, which gives a better chance to correctly classify a point. The increasing error when K > 45 can be justified by the fact that when the number of neighbors is more than a specific value, far points of the other class than the required point are included in the judgment process which increases the error probability.

3) After adding 4 new noise features to the original data set, the k with minimal average error was equal to 25 as shown in Figure 3 below.

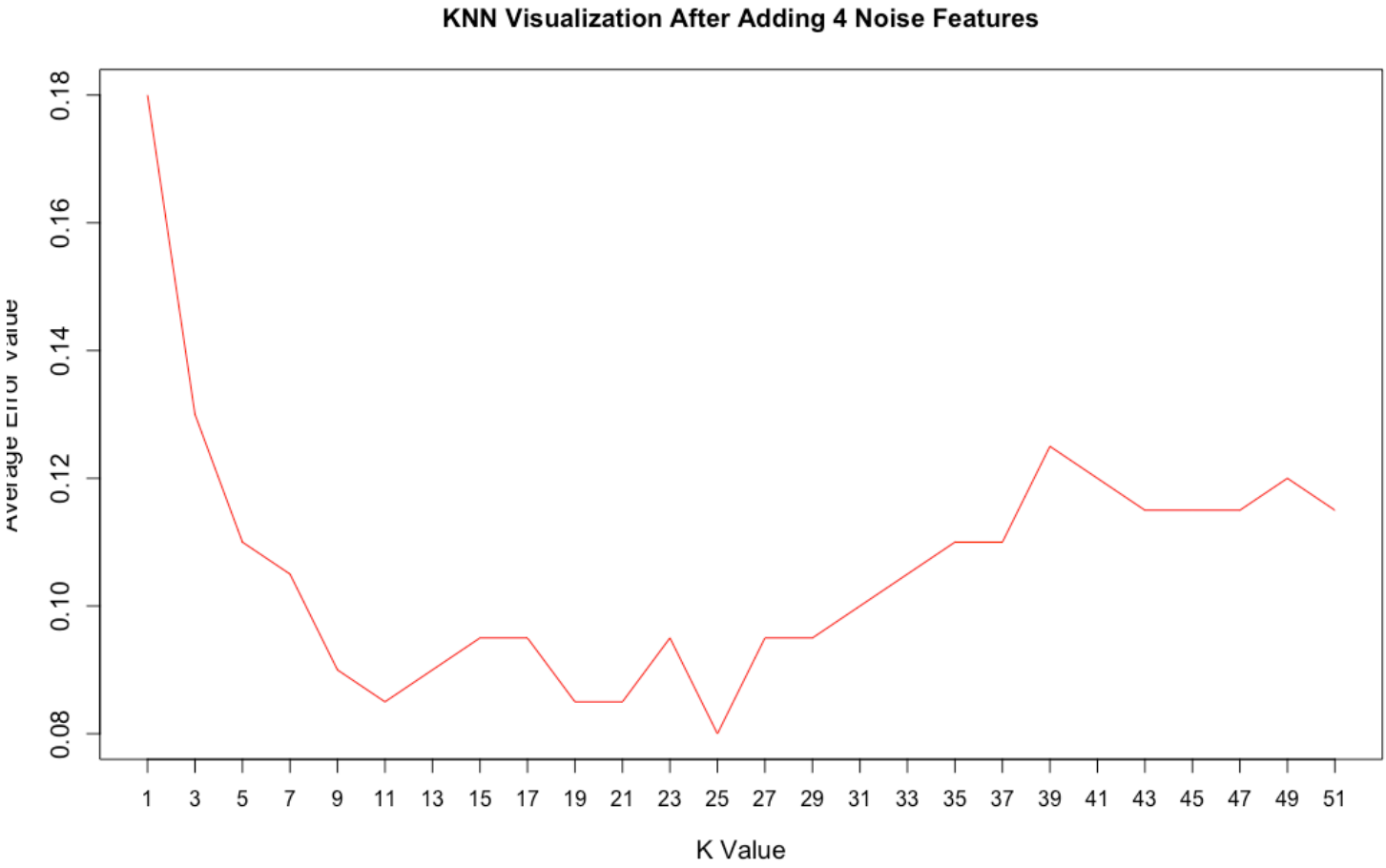


Figure 5: K vs Average Error After Adding 4 Noise Features

Explanation:

When noise features are added, the classifier tries to include the new features in the classification process. Which means the classifier will start from the noise in the data. Since the added noise is randomly generated, it will be hard to associate a class to the added features. This also means different classes might have close or similar values for the same noise feature. This explains the increasing error at k=1 because in this case classification will be done according to the nearest neighbor which might be of another class, however the reason for this classification can be because the noise features make these two points more alike compared to looking at the real data without the noise. The same also applies for other low k values.

K=25 shows the minimum error as judgment in this case will be done according to the nearest 25 points. In this case noise will not have the same effect as k=1. The reason behind that is the two original features still have some effect on the

classification process. This means point of the same class still have some features in common that can be discriminated from other classes.