# Question 1:

The result obtained after plotting the kNN classifiers using the Euclidean distance metric is as follow:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **K** | **Training error** | **Testing Error** | **Fitting** | **Variance** | **Bias** |
| 1 | 0 | 0.22 | Overfitting | High | Low |
| 3 | 0.143 | 0.205 | Overfitting | High | Low |
| 5 | 0.156 | 0.17 | Fitting |  |  |
| 10 | 0.152 | 0.17 | Fitting |  |  |
| 20 | 0.169 | 0.17 | Fitting |  |  |
| 30 | 0.169 | 0.16 | Fitting |  |  |
| 50 | 0.158 | 0.19 | Underfitting | Low | High |
| 100 | 0.196 | 0.2 | Underfitting | Low | High |
| 150 | 0.192 | 0.19 | Underfitting | Low | High |
| 200 | 0.222 | 0.205 | Underfitting | Low | High |
| 30 | 0.165 | 0.165 | Underfitting | Low | High |

The kNN classification with k=5,10,20 and 30 is giving us the low training error as well as low testing error. So, we can say that they are giving performance close to the fitting model. Among these classifications, k =30 can be considered the best classification as it is giving us the lowest testing error rate i.e., 0.16.

However, when setting k=1 or 3 the training error is very low, but the testing error rate is high. These two classifications performance can be considered as an Overfitting as they have high capacity. In terms of bias and variance we can say that they have high variance and low bias.

On the other hand, with k more than 50 the performance starts to decrease as the capacity is reduced. The kNN classification performance for k above 50 is giving Underfitting performance with low variance and high bias.

# Question 2:

The kNN classification using Euclidean distance metric with k=30 was giving us the best performance in the previous question with Training Error Rate at 0.169 and Test Error Rate at 0.16.

And, after plotting the kNN classification using Manhattan distance with k=30 we get the performance with training error rate at 0.165 and test error rate at 0.165.

So after observing these training error and test error, we can conclude that when using k=30, the kNN classification using Euclidean distance metric is giving us better performance in comparison to the kNN classification using Manhattan distance metric.