## K – NEAREST NEIGHBOR FOR CIFAR – 10 DATASETS

**a) Apply K-Nearest Neighbor algorithm with k = 1 on the test samples. We define the classification error rate as. What is the error rate of your classification?**

On using K =1, the error rate of classification was obtained as 0.73 for a sample of 100 test images where 73 samples were misclassified. Running for 200 test images gave an error rate of 0.7 with 140 samples misclassified. Similarly, the algorithm was tested for few difference test image length samples and results were tabulated

| S.NO | TEST IMAGES | ERROR RATE | MISCLASIFIED SAMPLES |
|---|---|---|---|
| 1 | 100 | 0.73 | 73 |
| 2 | 200 | 0.70 | 140 |
| 3 | 300 | 0.7366 | 221 |
| 4 | 400 | 0.7275 | 291 |
| 5 | 500 | 0.738 | 369 |

The plot of error rate for the above data results as below:

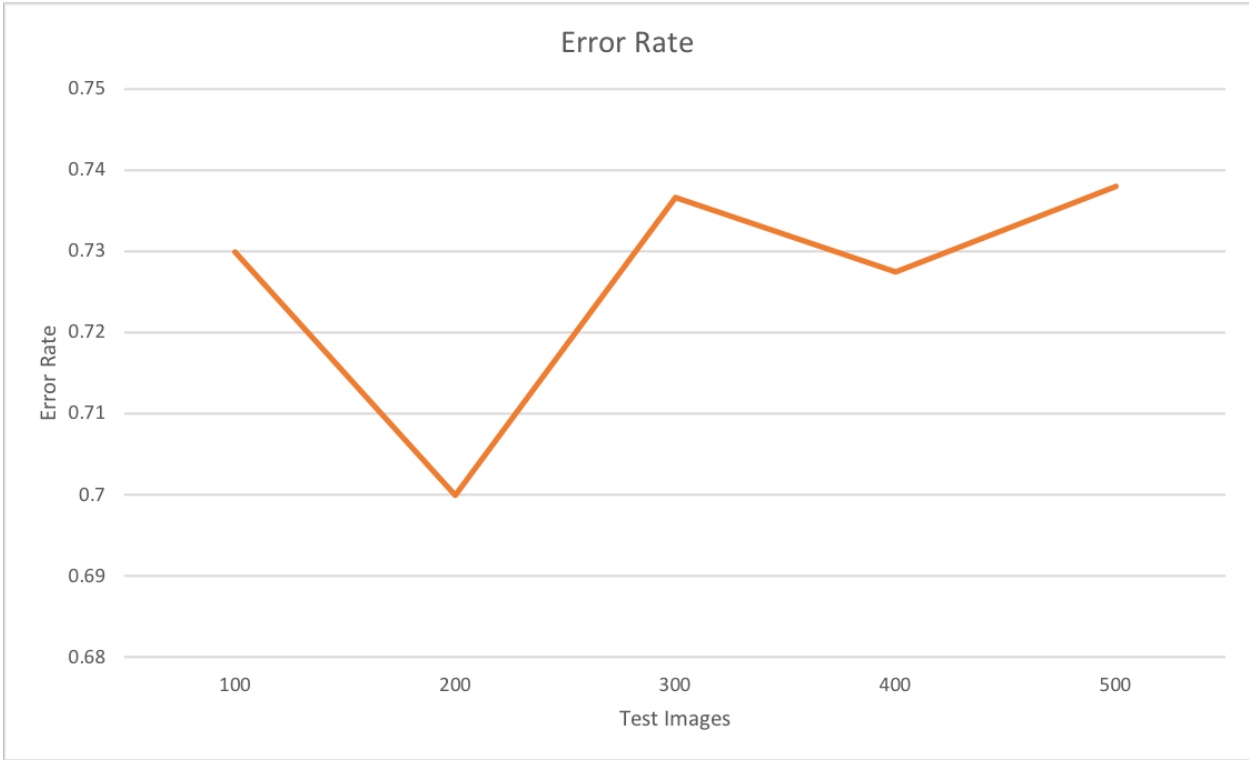


Figure 1: Error rate for K = 1 for various test samples

The misclassified samples were obtained by finding the number of indexes of the False values in predicted output

b) **Repeat last step for k = 2, 5, 10, 20 and plot the error rate $P_e$ against k. Is the error rate decreases with k? Should the error rate always decrease with k**?
For other K values, the error rate was again calculated for 100 and 200 test images and the values were tabulated

For 100 test images

| K VALUE | ERROR RATE | MISCALSSIFIED SAMPLES |
|---|---|---|
| 2 | 0.83 | 83 |
| 5 | 0.8 | 80 |
| 10 | 0.79 | 79 |
| 20 | 0.79 | 79 |

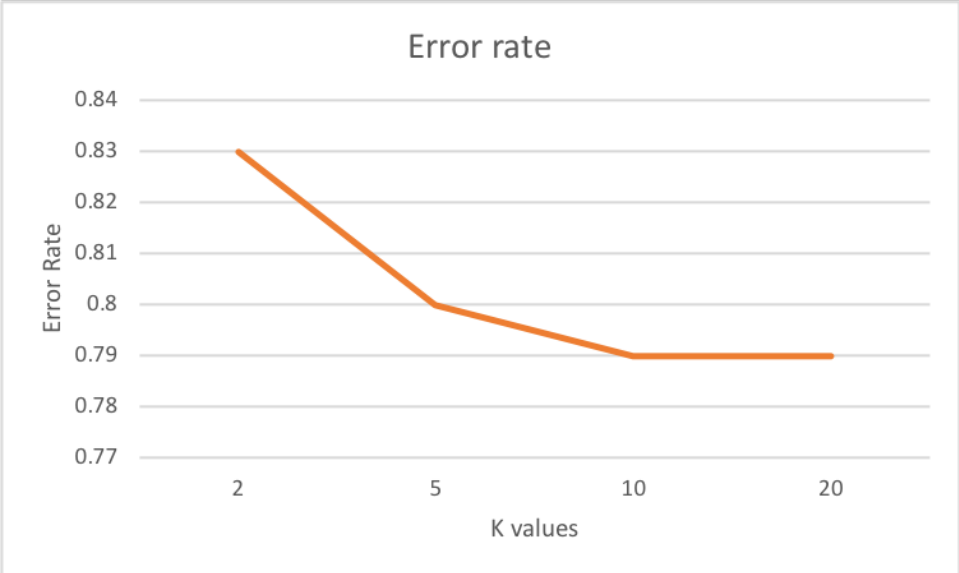


Figure 2: Error rate for various K for various 100 test samples

For 200 test images

| K VALUE | ERROR RATE | MISCALSSIFIED SAMPLES |
|---|---|---|
| 2 | 0.805 | 161 |
| 5 | 0.79 | 158 |
| 10 | 0.78 | 156 |
| 20 | 0.765 | 153 |

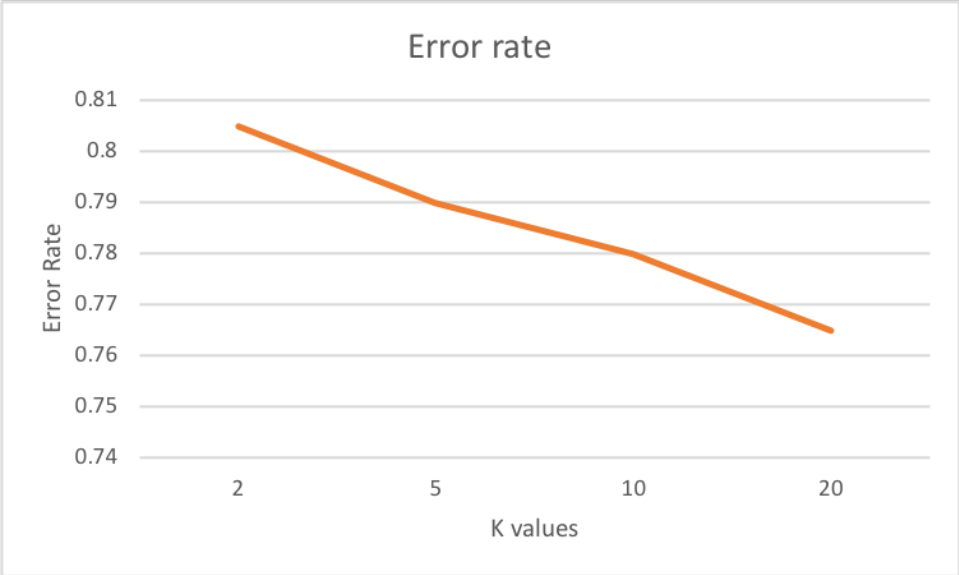


Figure 3: Error rate for various K for various 200 test samples

From the above 2 graph it can be seen that as K increases, error rate decreases but not by very large difference. Similar to this we can try for various test sample sizes and prove that as K increases, error rate decreases

**c) For each of the ten classes, pick a random image from test data and report its 10 nearest neighbors**

Let's consider K =20 and 200 test images and get the nearest possible neighbors for the misclassified samples based on distance using the predict function which gives the predicted label and the neighbors

0: [3, 0, 0, 8, 5, 8, 0, 5, 0, 0, 3, 8, 3, 8, 0, 9, 6, 3, 1, 8]
1: [3, 5, 3, 4, 0, 1, 7, 0, 0, 1, 1, 1, 9, 1, 4, 3, 1, 3, 3, 3]
2: [2, 8, 0, 2, 4, 4, 8, 2, 8, 4, 0, 2, 2, 0, 0, 2, 2, 8, 8, 0]
3: [5, 6, 4, 3, 5, 0, 3, 3, 3, 5, 4, 0, 4, 2, 1, 2, 3, 2, 5, 9]
4: [4, 4, 3, 7, 4, 4, 0, 4, 8, 4, 0, 4, 0, 6, 4, 3, 2, 1, 2, 8]
5: [3, 3, 3, 5, 7, 6, 5, 2, 5, 6, 8, 5, 6, 0, 2, 3, 6, 6, 5, 5]
6: [2, 6, 3, 2, 8, 6, 3, 6, 6, 4, 5, 6, 5, 2, 5, 6, 3, 5, 9, 3]
7: [7, 6, 2, 7, 8, 0, 7, 4, 7, 8, 4, 7, 3, 0, 7, 7, 0, 3, 1, 9]
8: [8, 8, 1, 7, 0, 1, 2, 4, 6, 3, 8, 1, 5, 0, 8, 2, 7, 3, 7, 0]
9: [9, 7, 9, 9, 5, 9, 1, 7, 0, 0, 7, 0, 1, 9, 9, 0, 8, 3, 9, 1]

The similar kind of findings were done for K = 10 and the observations were
0: [0, 3, 0, 8, 1, 0, 3, 0, 8, 0]
1: [0, 3, 8, 1, 2, 1, 7, 7, 8, 5]
2: [8, 8, 2, 0, 5, 1, 8, 0, 2, 2]
3: [4, 6, 1, 3, 5, 5, 5, 2, 3, 3]
4: [4, 2, 4, 4, 4, 4, 7, 0, 4, 4]
5: [5, 6, 6, 3, 5, 5, 2, 3, 5, 3]
6: [3, 6, 6, 2, 6, 6, 5, 3, 2, 5]
7: [6, 0, 7, 0, 1, 8, 3, 7, 7, 7]
8: [1, 0, 3, 8, 0, 8, 9, 8, 9, 9]
9: [0, 9, 1, 9, 9, 9, 0, 9, 9, 0]

**Reference:** https://gist.github.com/JeongUkJae/2725e018a3d06cccf1a8d3bf8c14c477. The template for executing KNN was obtained from this and made changes according to the question requirement