Report for Q5.

Testing Errors.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | s=5 | s=10 | s=50 | s=100 | s=500 | s=1000 |  |  |  | original |  |  |  |  |
| k=5 | 0.7777 | 0.7007 | 0.6529 | 0.6712 | 0.7023 | 0.7065 |  |  |  | 0.7066 |  |  |  |  |
| k=10 | 0.748 | 0.6789 | 0.6489 | 0.6684 | 0.7038 | 0.7097 |  |  |  | 0.7096 |  |  |  |  |
| k=50 | 0.7145 | 0.6489 | 0.6615 | 0.6863 | 0.7176 | 0.7198 |  |  |  | 0.7198 |  |  |  |  |
| k=100 | 0.7113 | 0.6592 | 0.6765 | 0.6969 | 0.7227 | 0.7249 |  |  |  | 0.725 |  |  |  |  |
| k=200 | 0.7158 | 0.6703 | 0.6944 | 0.7159 | 0.7385 | 0.7399 |  |  |  | 0.7399 |  |  |  |  |

Training Errors.

The train set size was really big for the computer to do the computations, so we chose 1000 first values in the train set to calculate the train error.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | s=5 | s=10 | s=50 | s=100 | s=500 | s=1000 |  |  | Original |  |  |  |  |
| k=5 | 0.545 | 0.459 | 0.471 | 0.482 | 0.544 | 0.548 |  |  | 0.547 |  |  |  |  |
| k=10 | 0.605 | 0.52 | 0.547 | 0.553 | 0.599 | 0.607 |  |  | 0.607 |  |  |  |  |
| k=50 | 0.661 | 0.6 | 0.621 | 0.644 | 0.684 | 0.678 |  |  | 0.678 |  |  |  |  |
| k=100 | 0.681 | 0.654 | 0.654 | 0.696 | 0.713 | 0.712 |  |  | 0.712 |  |  |  |  |
| k=200 | 0.693 | 0.696 | 0.696 | 0.71 | 0.741 | 0.742 |  |  | 0.742 |  |  |  |  |

From the tables that we get, we can see that for S=500 (less than half of the original dimension!), both training and testing errors are very close to the those of the original data, independent of the values of k.

To conclude, we think that PCA is effective for the KNN algorithm no matter what k we choose, as we can see that for any k, the error gets closer to the original one as s increases. Moreover for S=500 the errors are almost identical which means we succeeded to reduce the time of each calculation significantly (twice faster).