# Machine Learning - CSCI 5622

HW 1 - KNN

**Santhanakrishnan Ramani**

Friday 2nd September, 2016

## Analysis

1. Relationship between the number of training examples and accuracy.

| Data Limit | Accuracy |
|---|---|
| 10 | 0.2553 |
| 100 | 0.6700 |
| 500 | 0.8311 |
| 1000 | 0.8758 |
| 5000 | 0.9401 |
| 10000 | 0.9544 |
| 20000 | 0.9629 |
| 25000 | 0.9669 |
| 50000 | 0.9727 |

Table 1: Assuming K=3

Table 1 shows that the accuracy increases with number of training examples, although the increase in the accuracy with respect to the sample is high initially, after a certain point the increase is gradual.

2. Relationship between K and accuracy

| K | Accuracy |
|---|---|
| 1 | 0.9712 |
| 2 | 0.9721 |
| 3 | 0.9727 |
| 4 | 0.9739 |
| 5 | 0.9711 |
| 6 | 0.9733 |
| 7 | 0.9708 |
| 8 | 0.9724 |
| 32 | 0.9587 |
| 256 | 0.9170 |
| 1024 | 0.8462 |
| 16384 | 0.2026 |
| 32768 | 0.0764 |
| 50000 | 0.0983 |

Table 2: Trained using Full Dataset

Table 2 shows that accuracy increases a little initially till K value is 4, then it starts to decrease as K increases. A little wobbling can be seen between adjacent K values, which could be attributed to the median function used for classification as it outputs few values not in the label sets when K is even.

3. Numbers that get confused with each other most easily

Form the confusion matrix, it's evident that the numbers {2,7}, {4,9}, {5,3}, {5,6}, {8,3}, {8,5}, {9,7} gets confused easily which can be attributed to the structural similarity of these numbers while writing.