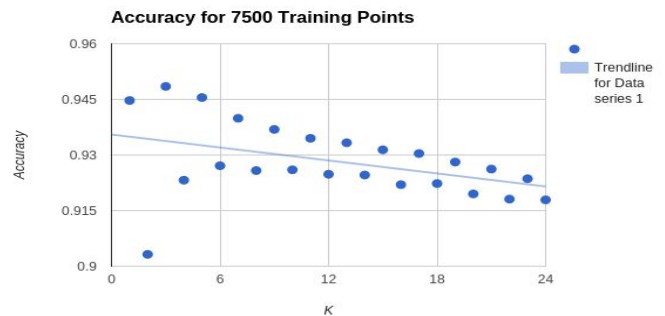
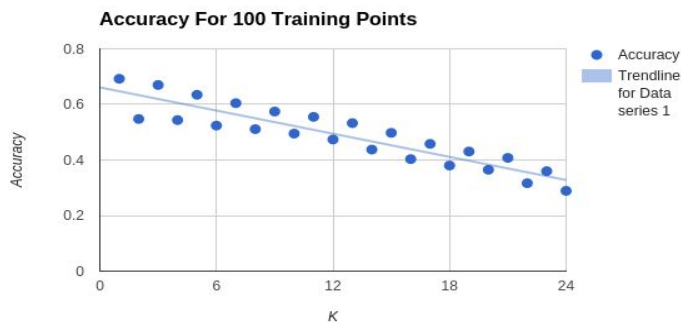


Andrey Shprengel  
CSCI 5622  
Homework 1

- 1) There is a positive correlation between the number of training point and the and the accuracy of the K Nearest neighbor algorithm for classification of our data set. The trend generally follows a logarithmic curve. Decisions on what size to use depend on computing power and the need for real time classification



- 2) The value of K generally has a negative correlation with accuracy. Odd values for K result in higher accuracy than neighboring even numbers. For small training sets the optimal k was 1 with a large decrease as K increased. For larger training data sets k=3 performed best with smaller change in accuracy for different K values. The following charts show the results



- 3) The numbers that confused most easily depended on the training data set but the most common ones were:

True Tabe	KNN Label
2	7
8	3
8	5

