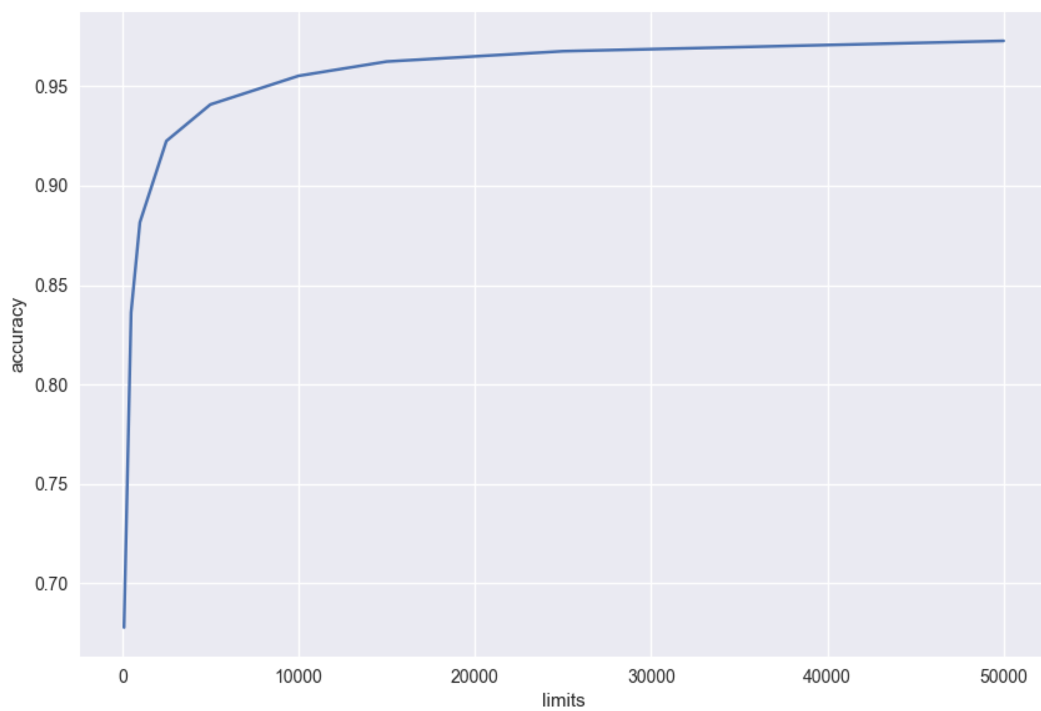
Machine Learning	**Homework 1**
CSCI 5622 Fall 2017	**Due Time Sep 15, 2017**
Name: Kumar Bhargav Srinivasan	**CU identitykey: kusr7198**

1 K-nearest Neighbor (40pts)

1.2 Analysis

1. What is the role of the number of training instances to accuracy (hint: try different -limit" and plot accuracy vs. number of training instances)?

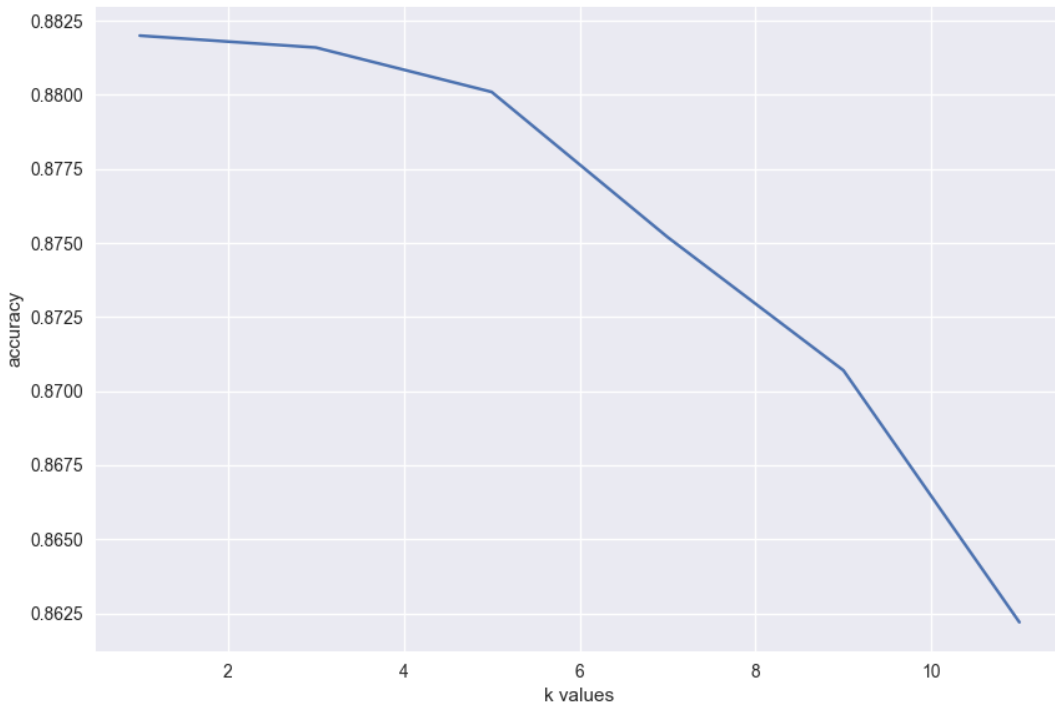
As the number of training instances increases, **the accuracy of the solution increases** (Error rate decreases)
Following is the graph which denotes different quantity of train data vs accuracy of the model.



2. What numbers get confused with each other most easily?
Following is the observations of the heat map.
Majority 1 : 4 and 9
Majority 2: 2 and 7

3. What is the role of k to training accuracy?

In general, As value of k increases above saturation, **accuracy decreases.**
Below is a graph which shows accuracy vs k values for training data set with 1000 observations, which denotes that if $k > 3$ then the accuracy decreases drastically.



4. In general, does a small value for k cause overtting or undertting?

From the analysis, we could say that K Nearest neighbors with k=1 implies to over-fitting, or in most of the cases would lead to over-fitting.

When k=1 the estimate of the probablity is based on the closest neighbor,a single point/sample. This is skews the decisions as there might be sort of distortions like outliers,mislabelling or noise in data. By higher value of K, decisions made will be robust to distortions.

2 Cross Validation (30pts)

2.2 Analysis

1. What is the best k chosen from 5-fold cross validation with -limit 500"?

best k chosen is 1 with accuracy 86.6 (results attached)

2. What is the best k chosen from 5-fold cross validation -limit 5000"?

best k chosen was 1 with accuracy of 93.8 (results attached)

best k will be pre-dominantly 3 for higher value of limits

3. Is the best k consistent with the best performance k in problem 1?

No, best k is higher than compared to problem 1 for limited dataset.

For Example : for k =3 with 5000 training data Accuracy in problem 1 is 93.1

whereas,

for k =3 with 5000 training data Accuracy in problem 2 is 93.82

3 Bias-variance tradeoff (20pts)

Assumptions :

1. For a given x, $Y(x) = f(x) + .$
2. $Variance(\in) = \sigma_{\in}^2$
3. $E(\in) = 0.$
4. $h_s(x_0) = \sum_{l=1}^{k} y_l$ where $x_{(l)}$ is the l'th nearest neighbor to x_0, implies $var(f(x_l)) = 0$
5. $Err(x_0) = E((y_0 - h_s(x_0)))^2$

We know from the bias variance tradeoff that $Err(x_0)$ is the following:
$E((y_0 - h_s(x_0)))^2 = bias^2 + variance + variance(\in)$
(As derieved by Chenhao Tan in Lecture)

Then the variance of this estimate is: $variance(h_s(x0)) = var((1/k) * \sum_{l=1}^{k} Y(x_{(l)}))$
$= (1/k^2) * \sum_{l=1}^{k} var(f(x_{(l)}) + \in_l)$
$= (1/k^2) * \sum_{l=1}^{k} (var(f(xi)) + var(\in_l))$
$= (1/k^2) * \sum_{l=1}^{k} var(\in_l)$, As var$(f(x_{(l)})) = 0$
$= (1/k^2) * (\sigma_{\in})^2 * \sum_{l=1}^{k}(1) = (1/k^2) * (\sigma_{\in})^2 * k$, According to assumption that $Var(\in) = \sigma^2$
$\boldsymbol{=}$ $(\sigma_{\in})^2/k$ **− (a)**

Bias is equal to difference between actual Target function and predicted function or hypothesis function which should be close in predicting target values.
Target function Y at some value x_0 is $f(x_0) + \in_0$
Square of the bias is as follows :
$bias^2 = (Y(x_0) - E_t(f_k(x_0)))^2$ where E_t is estimated target for function.
$=(Y(x_0) - E_t((1/k) * \sum_{l=1}^{k}(Y(x_l)))^2$
Assuming that $x_{(l)}$ is the l'th nearest neighbor to x_0, Expectation can be removed as it becomes actual prediction for x_l
$=(Y(x_0) - (1/k) * \sum_{l=1}^{k} Y(x_l))^2$
$=(f(x_0) + \in_0 - (1/k) * \sum_{l=1}^{k} f(x_l) + \in_l)^2$

Note : $\in_l$ is added to balance out the bias and the mean over all the values in x will be 0
$bias^2 = (f(x_0) - (1/k) * \sum_{l=1}^{k} f(x_l))^2 - -(b)$

Hence from (a), (b) and 2 ,
$E((y_0 - h_s(x_0)))^2 = bias^2 + variance + variance(\in)$ becomes,
$E(x_0) = (f(x_0) - (1/k) * \sum_{l=1}^{k} f(x_l))^2 + (\sigma_{\in})^2/k + \sigma_{\in}^2$