

# CMSC726: Assignment #1 — KNN

Soheil Behnezhad

September 8, 2017

## 1. What is the role of the number of training points to accuracy?

Basically, the higher the number of training points is, the higher is the accuracy. The dependence is not linear though. Increasing the number of data points when the number of data points is relatively low, has a higher impact on the accuracy.

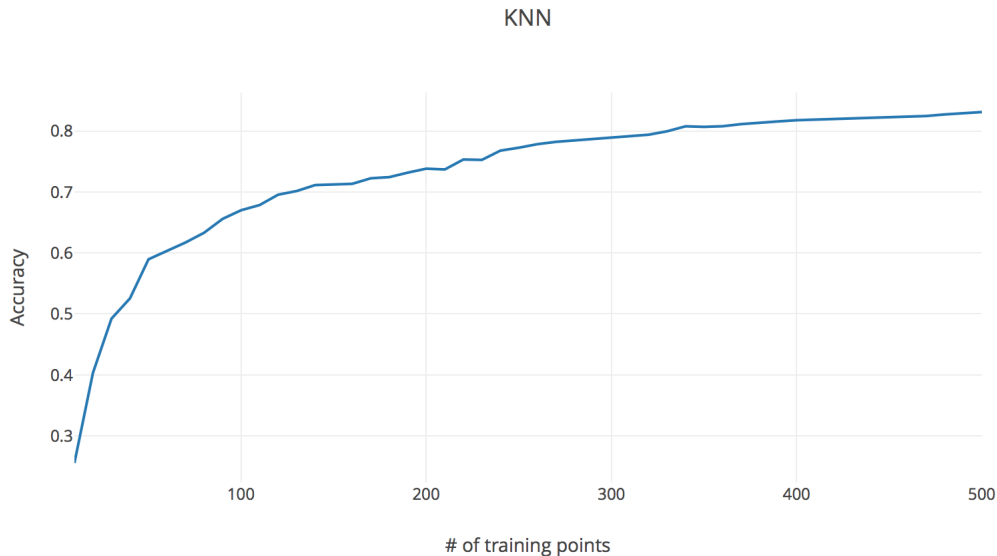


Figure 1: Accuracy based on the number of training points, I set  $k = 3$  here.

## 2. What is the role of $k$ to accuracy?

Generally, I had two observations (Figure 2 and 3). First, the parity of  $k$  is very important. That is, for odd values of  $k$ , the accuracy is generally higher than the even values. This makes sense since we use median in case of a tie, which is not a very wise thing to do. Second, small values of  $k$  seem to work better.

## 3. What numbers get confused with each other most easily?

The heat-map in Figure 4 visualizes the confusion matrix. For clarity, I just removed the main diagonal of the matrix. Our classifier is most confused with number 4 and guesses it to be 9. It also quite often guesses 7, 5, and 2 to be 9, 3 and 1 respectively. I also find it interesting that the reverse is not necessarily true, e.g., it doesn't guess 9 to be 4 that much.

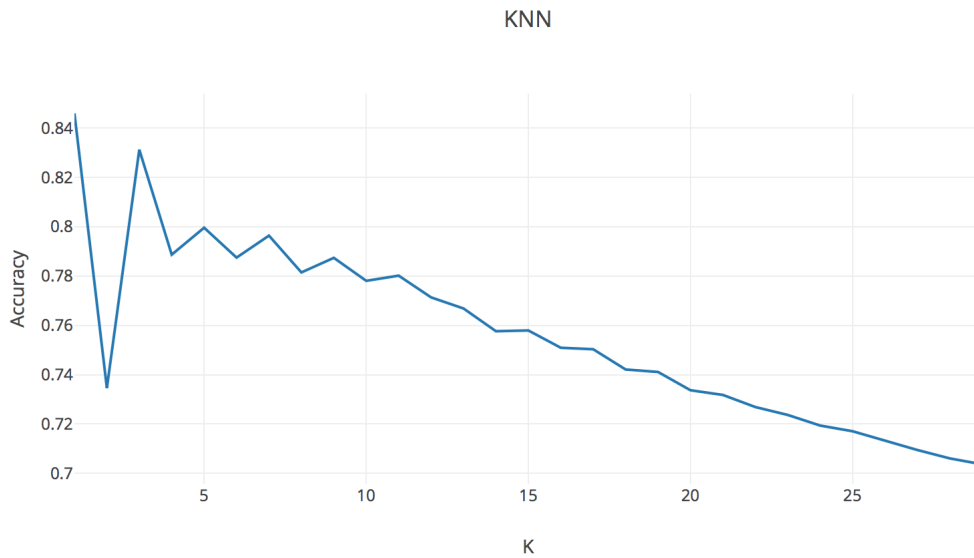


Figure 2: The importance of the parity of  $k$  in accuracy is clearly illustrated in the plot. The number of data points here is 500.

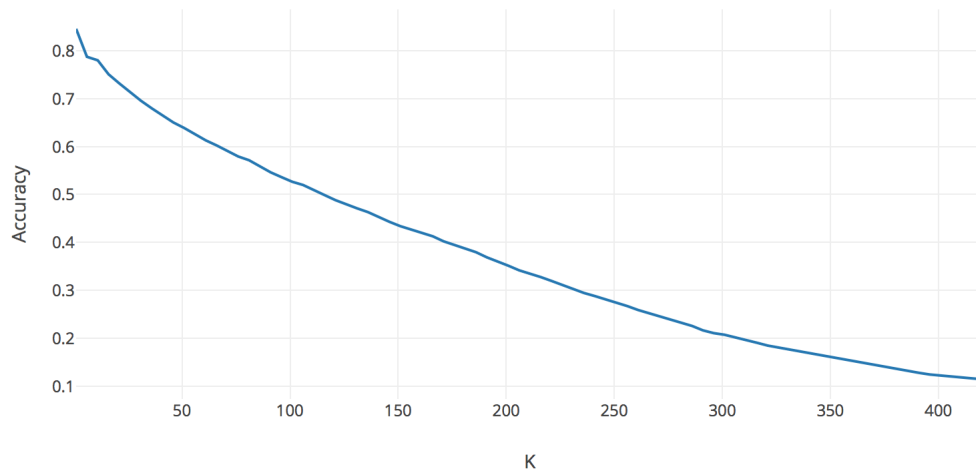


Figure 3: The same figure as Figure 2, which also includes larger values of  $k$  up to 450. Roughly speaking, the higher  $k$  is, the lower is the accuracy. The number of data points here is 500.

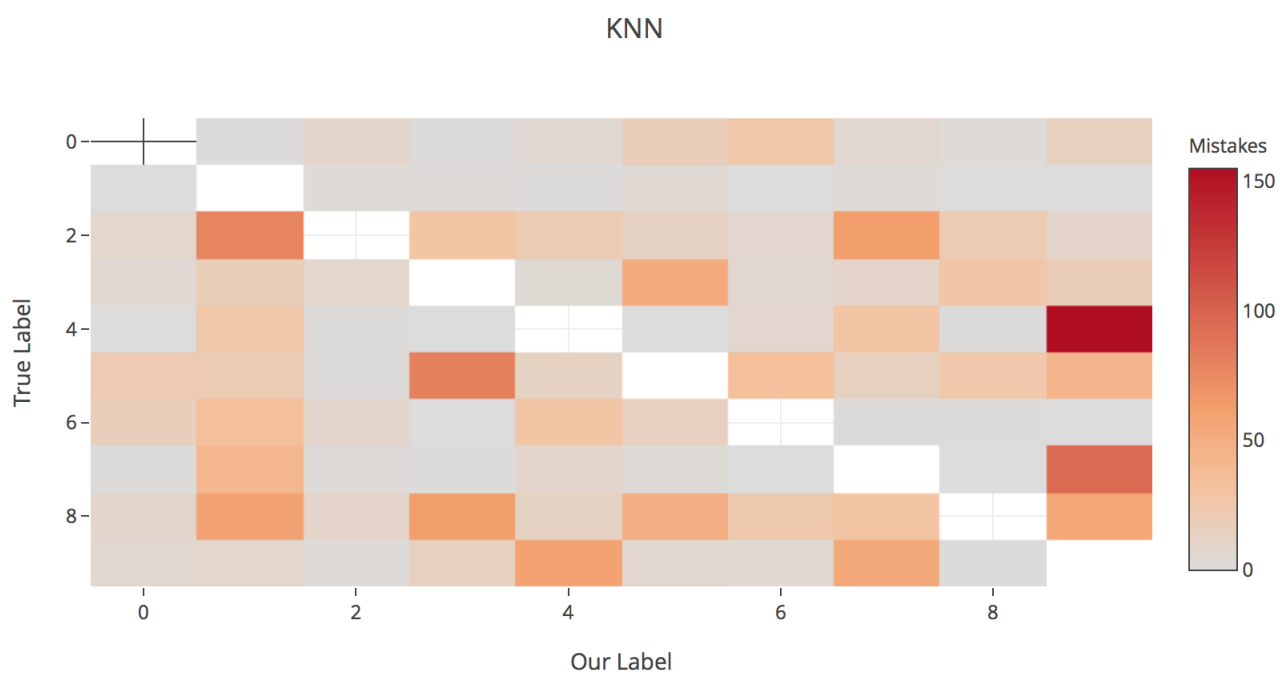


Figure 4: The confusion matrix visualized in a heat-map. The main diagonal was removed to make comparisons easier. The number of data points here is 500 and  $k$  is 3.