Sigurður Ágúst Jakobsson

Project: Assignment 2 - K-Nearest Neighbours

B. Theoretical

What explains this difference in accuracy between kNN and wkNN when k increases?

When increasing k in kNN we increase the likelihood of taking in targets from classes that do not match our actual class. Our 'circle' always gets larger, and depending on class distributions and variances, we always take in more targets that do not match our target in other words. This increases the likelihood of a majority of non-matching targets getting into our neighbor set. At the extreme we look at the whole dataset and are practically picking the most prevalent target in the data set, not the one that is in the middle of similar targets. kNN doesn't punish these distant votes.

In wkNN we account for the fact that the higher of a k we have the further many of the neighbors will be away. It detracts the weight of faraway neighbors, so we are still focusing on neighbors that are close to our test value.