KNN RESEARCH REPORT

Choosing a value for k

Abstract

[Draw your reader in with an engaging abstract. It is typically a short summary of the document.
When you're ready to add your content, just click here and start typing.]

Chris Dilger
[Email address]

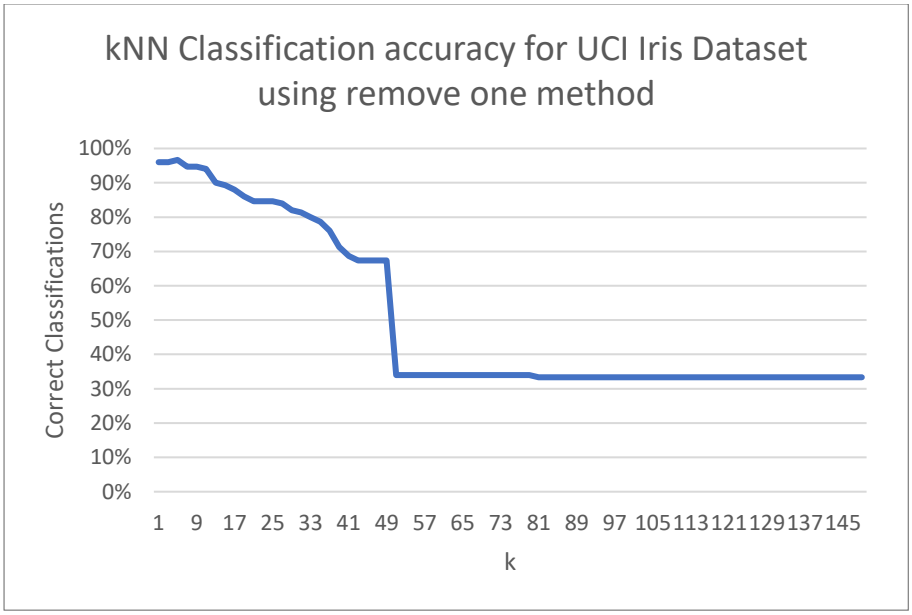
Research question

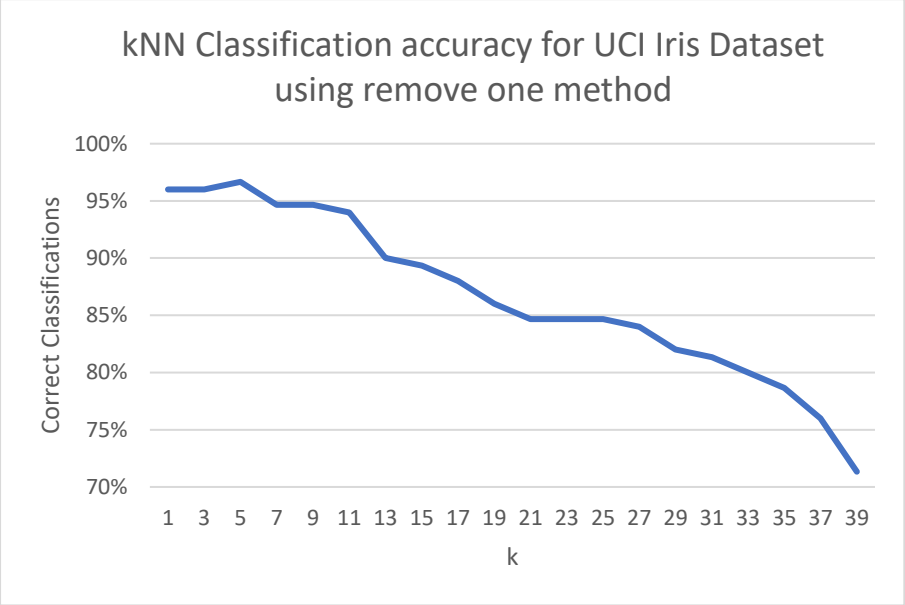
For which value of k is the highest accuracy of classification achieved?

Method

Remove one datapoint, construct a training dataset from the rest and compare the knn_search classification with the actual point classification. Divide the number of correct classifications over the total number of points. Repeat for each value of k.

Results





We see approximately a 95% accuracy for k values 1-9, with the best accuracy at k=5.

Discussion

- Dimensionality
- Generalizations to be made
- More data

- Other datasets
- The dropoff at 1/3
- What might happen if there are too many data points of one category
- When normalization might be useful