CMSC630-A3

Bobby Best

May 2019

1 Features

The features used for the KNN model are the following:

1. The number of 'object' pixels of the segmented images (one for each color channel + the grayscale image)
2. The standard deviation of the grayscale histogram
3. The highest peak in the grayscale histogram
4. The lowest valley in the grayscale histogram

2 K-Nearest-Neighbors Model

The KNN model takes in two datasets of the features described above, one with labels (train) and one without (test), and it assigns to each test sample the value of the most frequently occurring class in the k nearest training samples to it.

3 Evaluation

The evaluation step uses the predicted labels from the KNN algorithm and the set of corresponding gold standard labels to compute micro-averaged precision, recall, F1 score, and accuracy.

4 Cross Validation

Cross validation is performed by taking in a dataset and its labels with an integer value 'cv'. This value represents the number of folds, i.e. the number of subsets to split the dataset into. Each subset is then iteratively used as the test set with the remaining being used for training, and at the end all of the results are averaged to produce the final metrics.

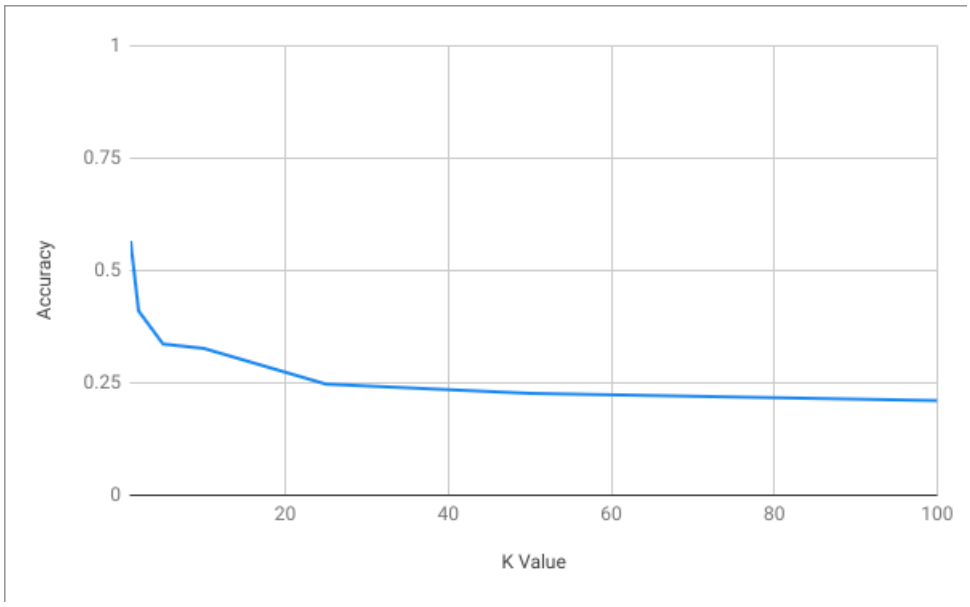

Figure 1: Plot of Accuracy with increasing values of K