Hannah Arntson – hra069
Peter Haddad – pbh423
Katie George – kmg381

Features

- Bounding Box area
- Curvature
- Bounding Box Height to Width Ratio

All of these features were discrete.

To determine the thresholds for the function, we created a function that calculates the value of the feature on all objects in the training set. It then returned two lists, one containing all of the feature values for text, and one for drawings. We then did some statistical analysis on these values to try and find the best split point. We calculated the mean, median, standard deviation and other values. We also used matplotlib to plot histograms of these features to visually compare the different values. An example histogram is below. It plots the frequency of area for text sketches. This, when compared to the drawing histogram, showed us that values below 10000 are most likely to be text, so we used that as a splitting point.
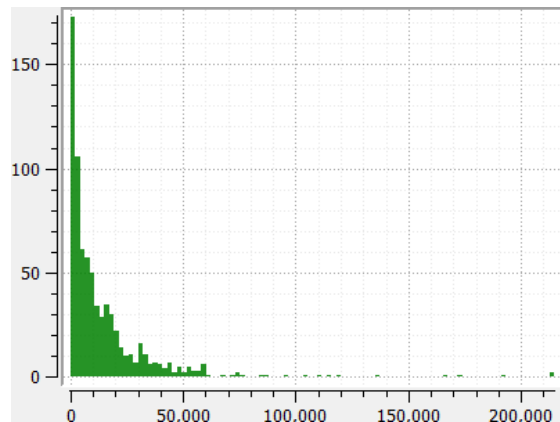
Confusion Matrix

Best

	Drawing	Text
Drawing	933	341
Text	298	446

Basic

	Drawing	Text
Drawing	855	419
Text	301	440