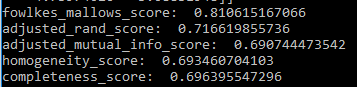
Given the set of data provided, our goal was to categorize the wheat into three different varieties. The dataset was comprised of 210 samples, each with seven attributes that described the wheat kernels. The true variety was also provided as an additional attribute in the dataset.

To execute K-Means clustering on the dataset, I used KMeans from the sklearn.clustering python library. I used several different scoring methods including the Fowlkes-Mallows index which yielded the highest indicated accuracy rating at 81%. The others were very close to each other and indicated about 70% accuracy as seen below. I tried several parameters within the KMeans constructor. I changed max\_iter to lower and much higher values, but that yielded no change. Next I adjusted n-init to higher values to get a larger variety of centroid seeds. This also yielded no change in scores. The final parameter I changed was init, so that it used ‘random’, but that did not change the results. Since none of those parameter changes improved the results I changed KMeans back to the defaults, except for the number of clusters.



At the end, KMeans resulted in the below centroid coordinates. The first two attributes seem to have the biggest influence and vary the most. The other attributes vary by less than one, attribute three varies by much less than one.

