Assignment 6

Philip Lassen

January 8, 2019

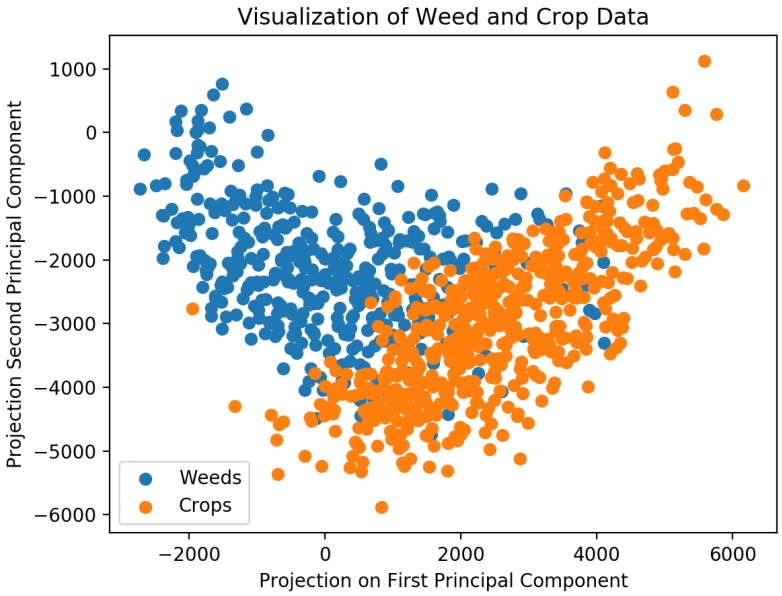
Contents

1 Unsupervised Learning – Clustering

The code can be found in the file main.py which is in the zip file.

Task 1: Visualisation of Input Data

I made a scatter plot for the data visualization utilizing the projection of the datapoints onto the two principal components of the data. The image can be seen below

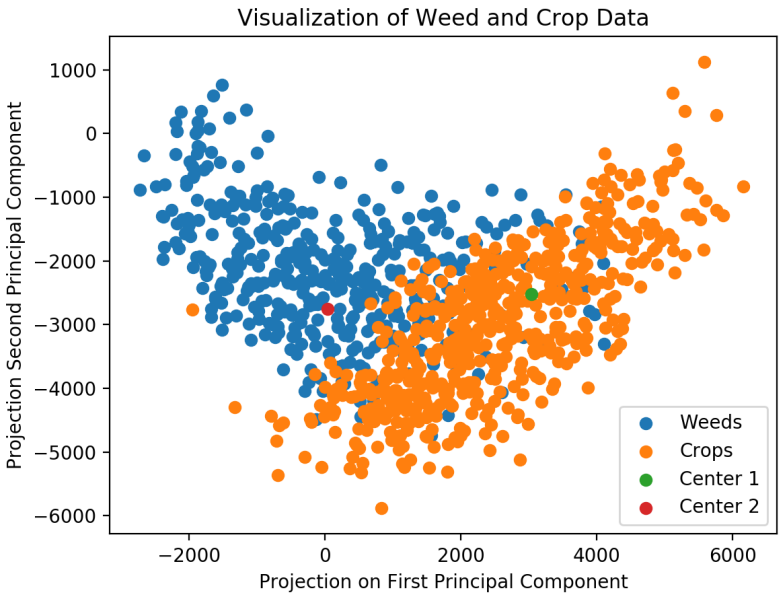


Task 2: Clustering

For the 2-Means Clustering I decided to use the SKLEARN library with their implementation of K-Means clustering. The library was very easy to used and the model could be generated from the following simple one line python call.

```
kmeans = KMeans(n_clusters = 2, init = train[0:2, 0:13]).fit(train)
```

Below is the scatterplot from Task 1 but now with the two centroids generated from Kmeans clustering's projections added as well.



From the visualization above, we can see that the 2-means clustering seems to do a good job when looking at how it compares with the projections of the first two principal components.