CS 422: Project 2 Write-up

Sophie Liu

October 2022

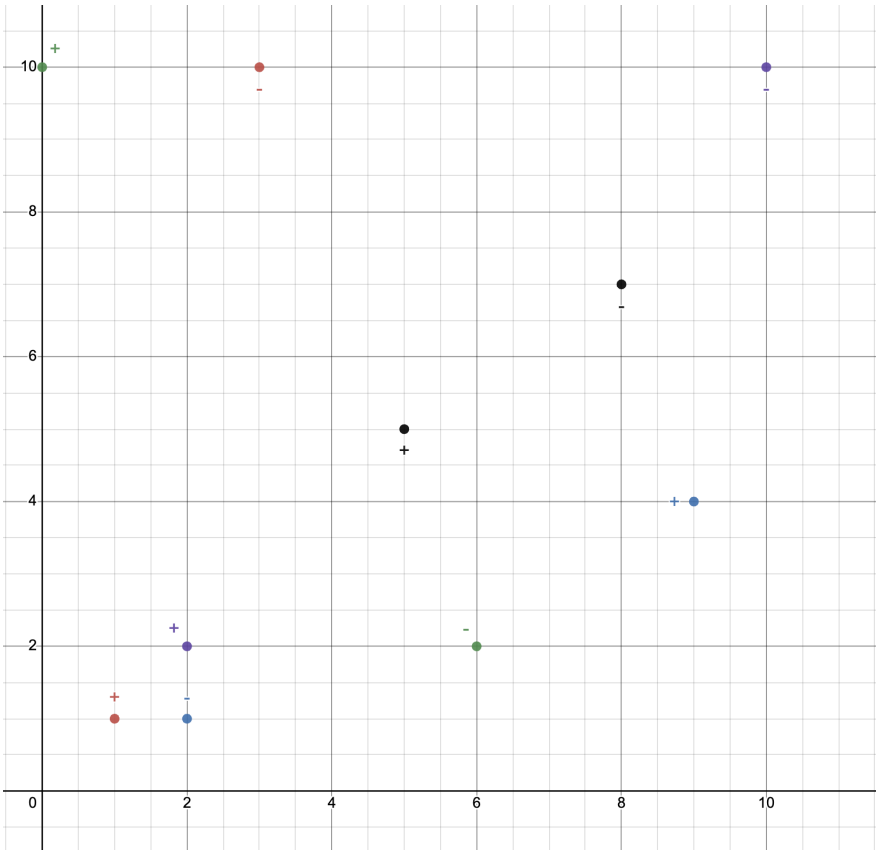
This is the link of my write-up: https://www.overleaf.com/read/cbknfnfjvqff.

1 Nearest Neighbors Implementation

Using the training data provided in "nearest neighbors 1.csv," my algorithm would classify the test points with K=1, K=3, and K=5 as follows:

Test points	K = 1	K = 3	K = 5
(1, 1, 1)	1	1	1
(2, 1, -1)	-1	1	1
(0, 10, 1)	1	1	1
(10, 10, -1)	-1	-1	-1
(5, 5, 1)	1	-1	1
(3, 10, -1)	-1	1	-1
(9, 4, 1)	1	-1	-1
(6, 2, -1)	-1	1	1
(2, 2, 1)	1	1	1
(8, 7, -1)	-1	-1	-1

The best K value for the training data above is difficult to say. Based on accuracies, we know that K = 1 has an accuracy of 10/10, K = 3 has an accuracy of 5/10, and K = 5 has an accuracy of 7/10. However, K = 1 would be overfitting. There's really not much of a large scale pattern in the points, either, and the difference in accuracy between K = 3 and K = 5 might just be luck. K = 5 might be a bit underfitting, so perhaps K = 3 would be better.



2 Clustering Implementation

The training data in "clustering 2.csv" with K=2 and K=3 result in some of the following KMeans.

It seems as if using KMeans on K = 2 seems to have larger variance in its cluster centers than KMeans on K = 3, suggesting that the structure of "clustering2.csv" has a more distinct structure as three clusters than as two clusters.

K = 2

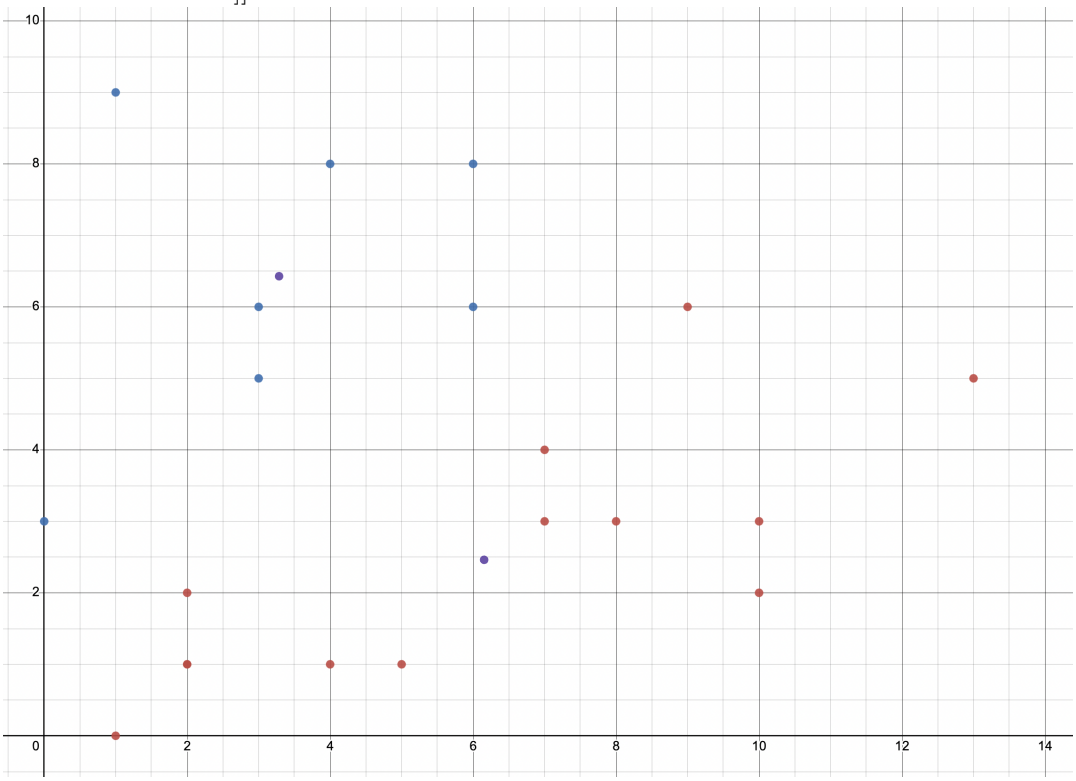
KMeans: [[3.28571429 6.42857143] [6.15384615 2.46153846]]
KMeans: [[2.45454545 3.36363636] [8.44444444 4.44444444]]
KMeans: [[8.44444444 4.44444444] [2.45454545 3.36363636]]
KMeans: [[8. 4.8] [2.3 2.9]]
KMeans: [[5.46153846 2.23076923] [4.57142857 6.85714286]]

K = 3

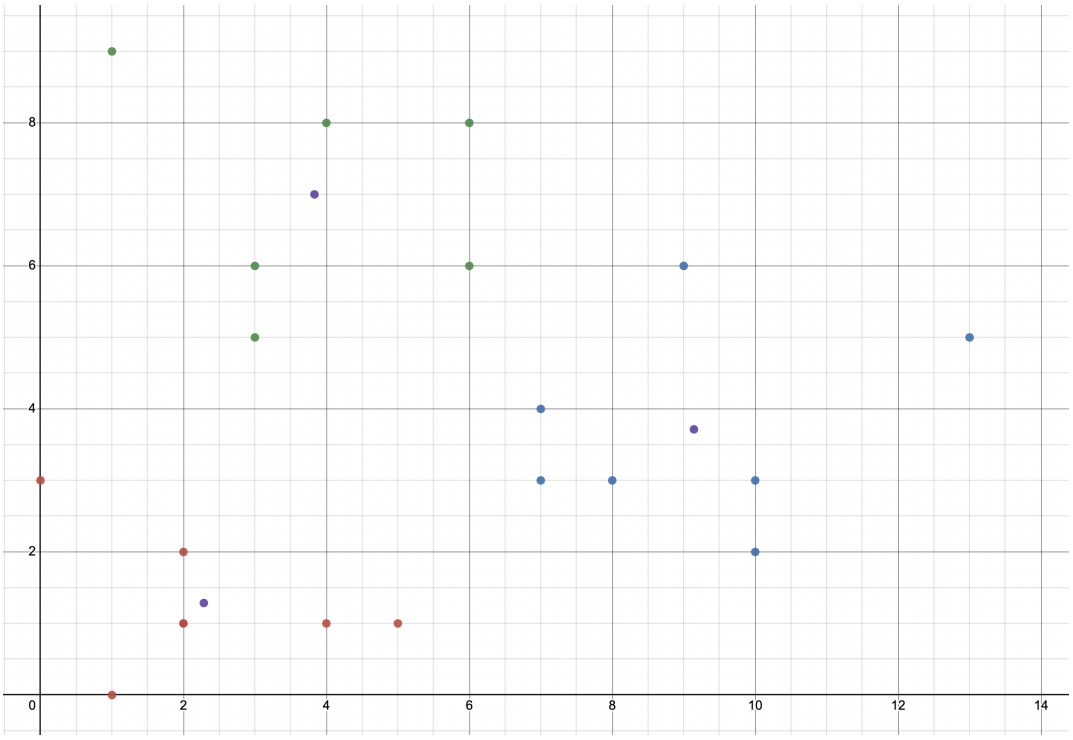
KMeans: [[2.28571429 1.28571429] [9.14285714 3.71428571] [3.83333333 7.]]
KMeans: [[9.14285714 3.71428571] [3.83333333 7.] [2.28571429 1.28571429]]
KMeans: [[3.83333333 7.] [9.14285714 3.71428571] [2.28571429 1.28571429]]
KMeans: [[9.14285714 3.71428571] [3.83333333 7.] [2.28571429 1.28571429]]
KMeans: [[10.5 4.] [2. 3.88888889] [6.14285714 3.71428571]]
KMeans: [[13. 5.] [2.45454545 3.36363636] [7.875 4.375]]
KMeans: [[11. 3.33333333] [6.71428571 5.42857143] [2.3 2.9]]
KMeans: [[2.28571429 1.28571429] [3.83333333 7.] [9.14285714 3.71428571]]

Graphing

A graph of the clusters with $K = 2$ and centers at [[3.28571429 6.42857143] [6.15384615 2.46153846]]:

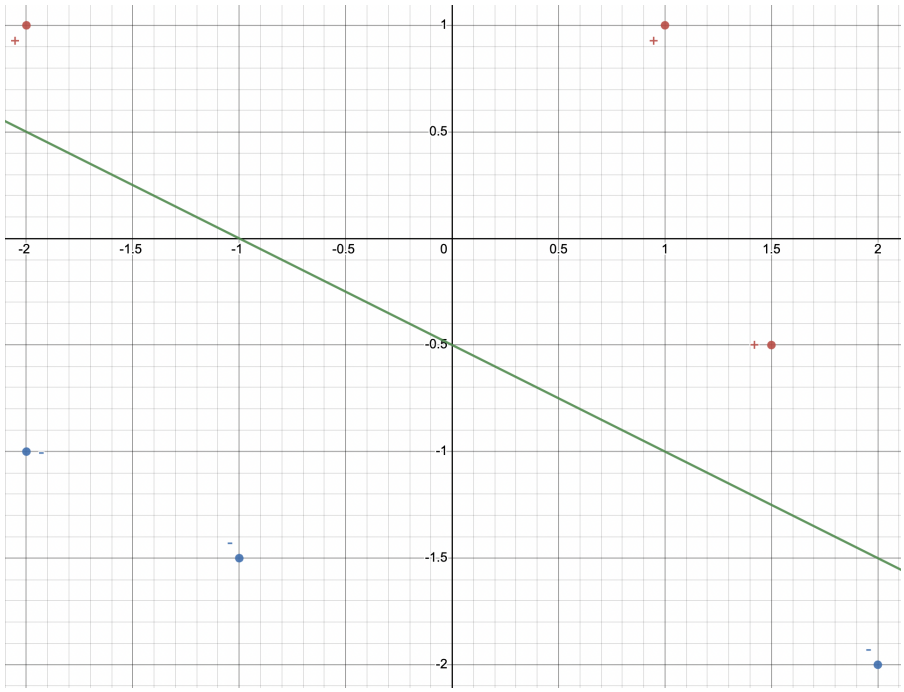


A graph of the clusters with $K = 3$ and centers at [[2.28571429 1.28571429] [9.14285714 3.71428571] [3.83333333 7.]]:



3 Perceptron Implementation

My perceptron resulting from training on the dataset provided in "perceptron 2.csv" yield $w =< 2, 4 >$ and $b = 2$. Below, the decision boundary is plotted along with the dataset.


1
0.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
-0.5
-1
-1.5
-2
+
+
+
-
-
-