Homework 4 Report
CSE 512
By Yasha Singh - 112970310

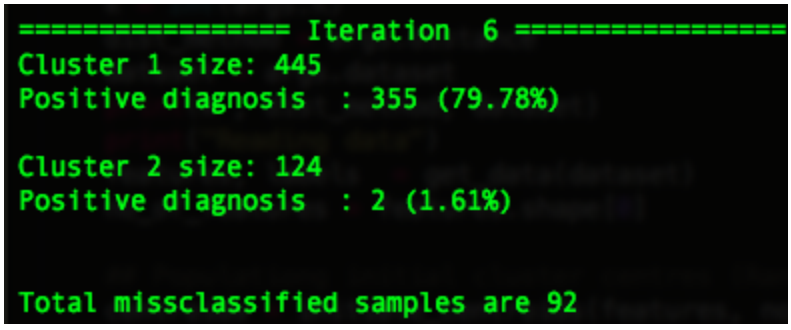
1. Observation with clustering K = 2
2. Results reported on Breast_cancer_data.csv after convergence in the final iteration of the algorithm.

	Cluster 1 size	Positive	%	Negative	Cluster 2 size	Positive	%	Negative
1st Run	445	355	79.78%	90	124	2	1.61	122
2nd Run	447	355	79.42%		122	2	1.61	120
3rd Run	445	355	78.79%		124	2	1.61	122

The results with both Euclidean and Manhattan distance were the same.

From the results we can say that K-means clustering was to quite an extent (with ~70 accuracy) actually able to group the positive diagnosed data points together in a single cluster. Especially because it is an unsupervised learning algorithm and it was still able to perform relatively well.

However, some of the patients with negative diagnosis also got wrongly clustered with the positive diagnosed patients. This is whether the algorithm suffered.



Total misclassifications - 92/569 = 16.168%
Accuracy = 83.83%