

DS-630: Project 1
Assigned Nov. 7, 2017
Due Nov. 21, 2017

In your first project you will investigate the data set *cluster.csv*. This data set consists of 1077 points, or observations, in a 70 dimensional feature space. Please note that the first column of the data set simply contains the number of the observation and is not a feature vector.

Your job is to determine the correct number of clusters for the data set. *There is* a ground truth to the data set, however as the researcher, you do not have access to this information. You are to code the EM algorithm in R as discussed in class to determine the correct number of clusters and cluster membership for each observation.