

CS 484

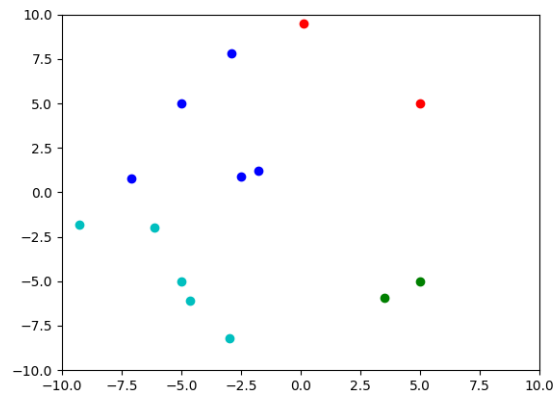
Assignment 2

Hieu Le

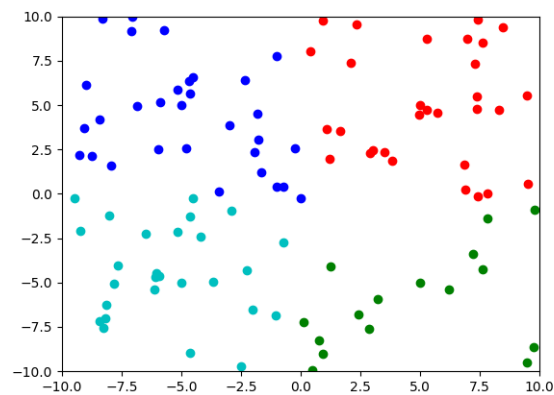
Run K-means with 4 categories

The initial centroids are positioned at $(5, 5)$, $(5, -5)$, $(-5, 5)$, $(-5, -5)$.

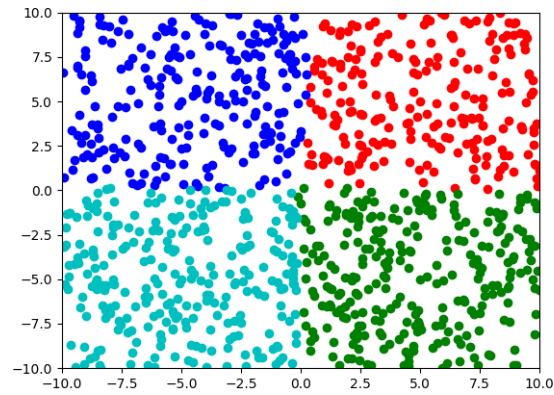
When the algorithm is run on a training set of 10 random observations (not including the initial centroids), the distribution of data points between the four categories are 2, 2, 5, 5. The positions of the final centroids end up at $(2.56, 7.24)$, $(4.26, -5.47)$, $(-3.85, 3.15)$, $(-5.60, -4.62)$.



For 100 random observations, the distribution are 30, 16, 31, 27. The positions of the final centroids are $(5.27, 4.90)$, $(4.68, -6.08)$, $(-4.74, 4.36)$, $(-5.53, -4.38)$.



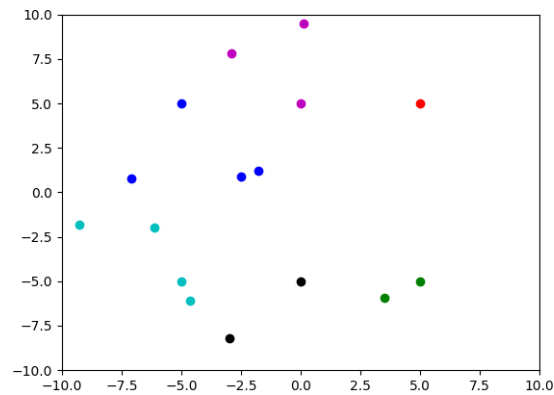
For 1000 random observations, the distribution are 223, 277, 235, 269. The positions of the final centroids are $(5.17, 5.03)$, $(4.76, -4.72)$, $(-4.51, 5.26)$, $(-5.01, -4.84)$.



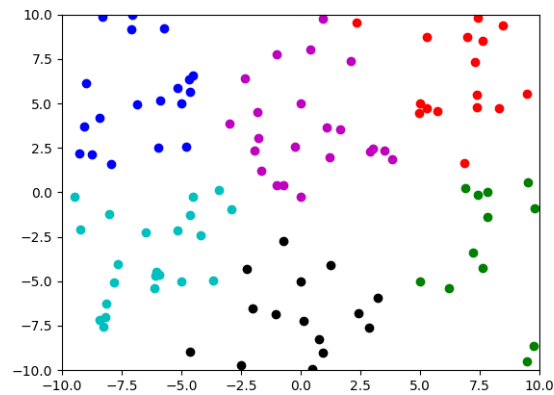
Run K-means with 6 categories

The initial centroids are positioned at $(5, 5)$, $(5, -5)$, $(-5, 5)$, $(-5, -5)$, $(0, 5)$, $(0, -5)$ with two additional means on the horizontal axis.

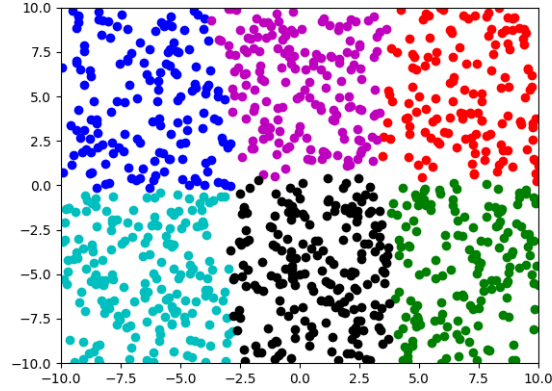
Partition 10 random observations into 6 categories:



Partition 100 random observations into 6 categories:



Partition 1000 random observations into 6 categories:



Run Learning Vector Quantization (LVQ) with 6 categories

The training data set is generated from a set of random data points from the domain. Each category contains an equal number of data points in the training set.

The initial weight matrix is initialized by taking one pattern from each category.

$$W = \begin{bmatrix} -5.0 & 0.0 \\ 1.0 & 7.0 \\ 10.0 & 3.0 \\ -9.0 & -3.0 \\ 1.0 & -6.0 \\ 7.0 & -1.0 \end{bmatrix}$$

The neural network is then trained by supplying the entire training data set together with their expected category. The trained network is then tested on a test set consisting of 1000 random data points.

The performance of the neural network seems to be improved by increasing the size of the training set. Below is a table denoting the number of correct categorizations for each training set size.

Training Set Size	Number of Correct Categorizations
6	659
12	792
18	757
24	909
30	859
36	877
42	928