CSCI E-181 Spring 2014 Practical 1

David Wihl
davidwihl@gmail.com

January 31, 2014

Warm-Up

Initially I used a K-Means implementation in Octave I had written for a previous course[1]. While this implementation was sufficient for the prior course's Dataset, when I tested it with five clusters of random data, K=5 and random initial centroids, one of the centroids would frequently not converge on any points.

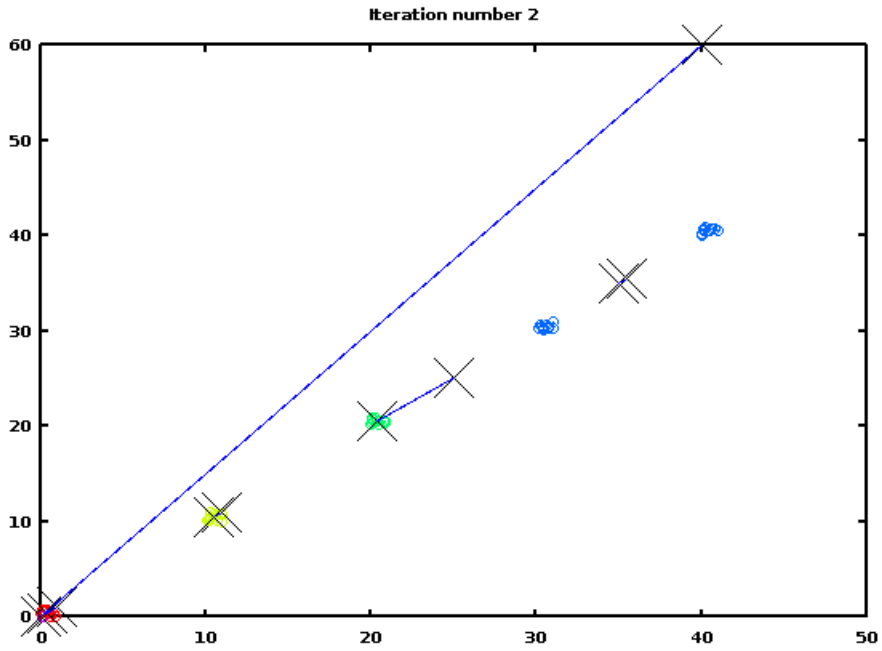

Figure 1: Random Initial Centroids After 1 Iteration

[1]Machine Learning, Coursera, Prof. Andrew Ng, Completed Jan 2014, https://class.coursera.org/ml-004

I subsequently modified the code to use K-Medoids, choosing one of the sample data points at random as an initial centroid. This worked much better.

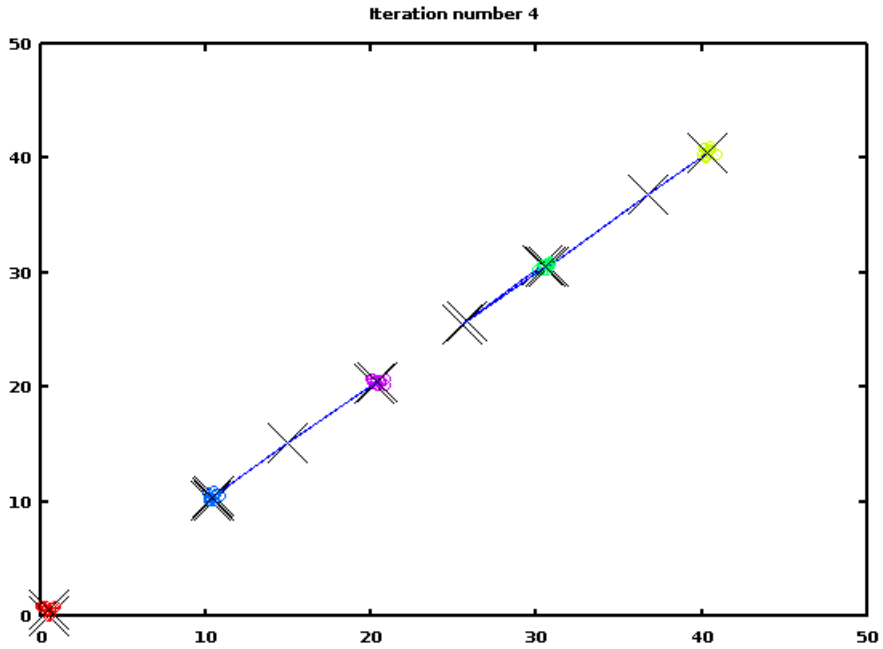

Figure 2: K-Medoids Converge After 4 Iterations

CIFAR-10 Image Data

I then attempted using K-Medoids with the CIFAR-10 Image Data, using the Matlab version with Octave. The training data consists of a 10000x3072 matrix of UInt8. Each row is a 32x32x3 (=3072) color image, consisting of 1024 red, 1024 green and 1024 blue elements. There are 10 classes in the set ("airplane", "automobile", etc.), so setting K=10 was a rational first step.