Homework 5

Geoff Li

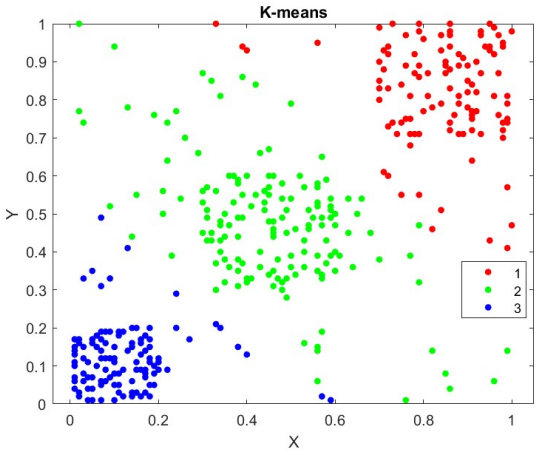
ligen4@msu.edu

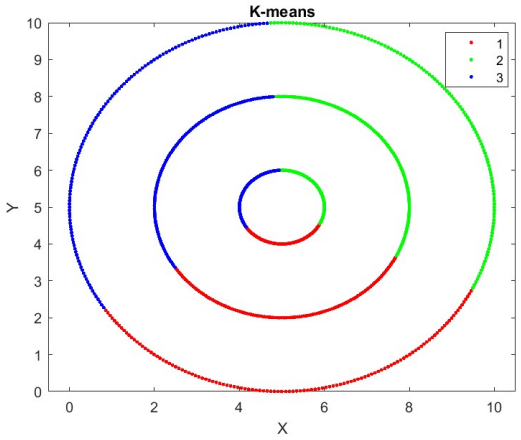
Clustering: k-means

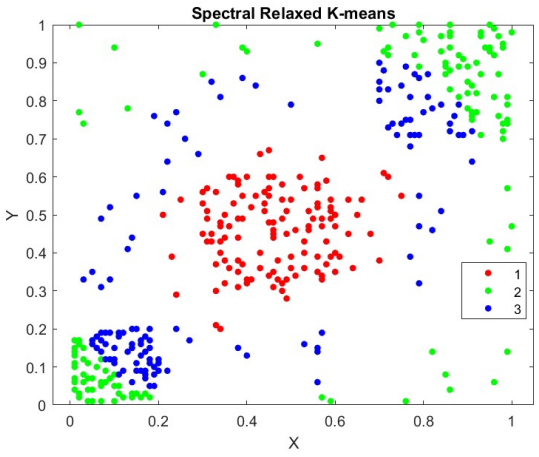
Question 1

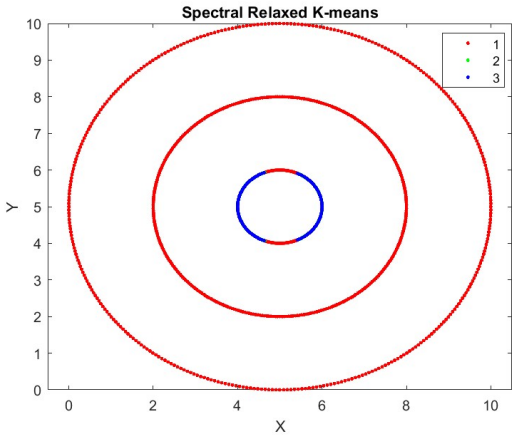
The spectral relaxed k-means adds pre-process procedures before using k-means. It transforms the original data using eigenvectors. It's possible that the spectral k-means gets the same solution with k-means if the data keeps the same after transformation.

Question 2





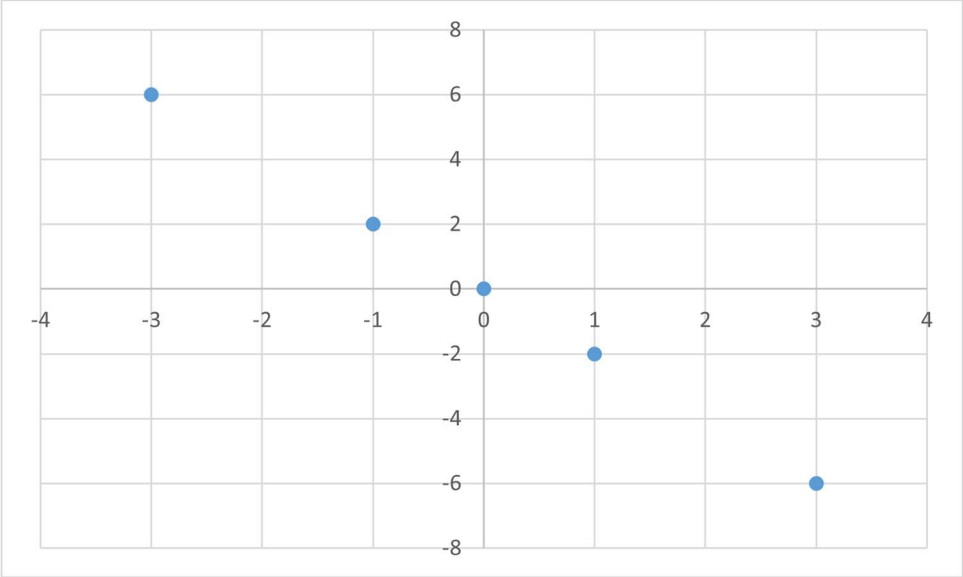




For the first random points dataset, it's obvious that k-means gets better clustering results.

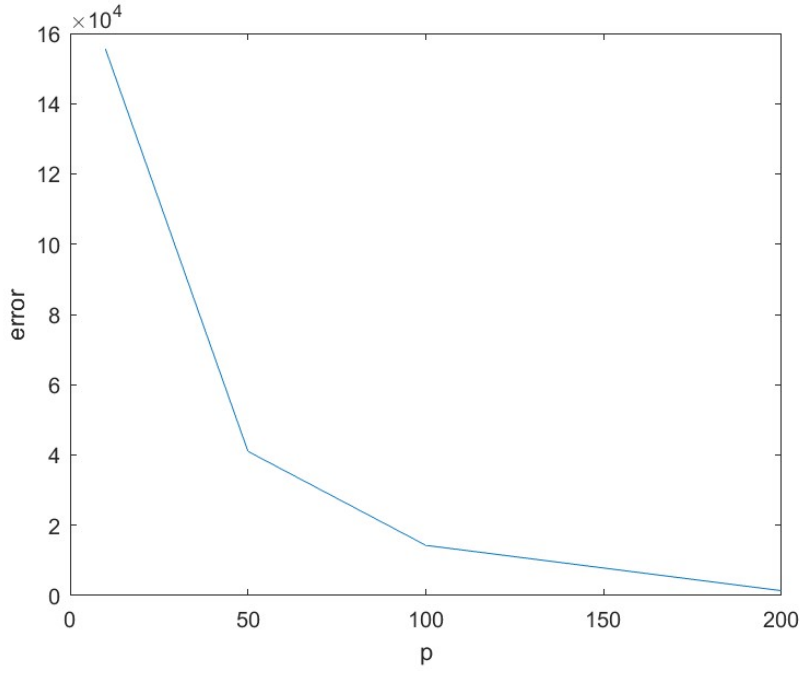
Principle Component Analysis

Question 1



The first principal component is the direction in space along which projections have the largest variance. The second principal component is the direction which maximizes variance among all directions orthogonal to the first. So the first principal component is $(\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}})$, and the second component is $(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$.

Experiment



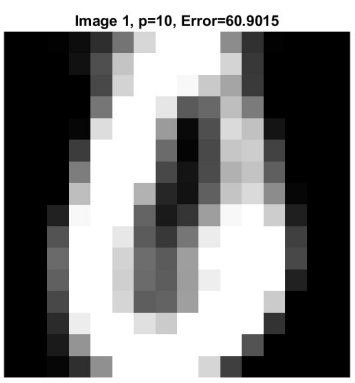


Image 1, p=10, Error=60.9015

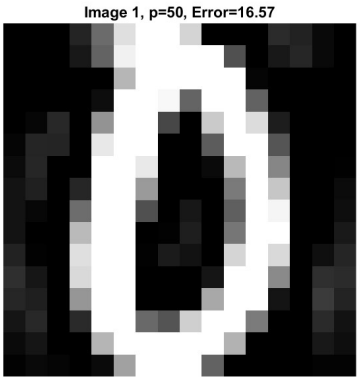


Image 1, p=50, Error=16.57

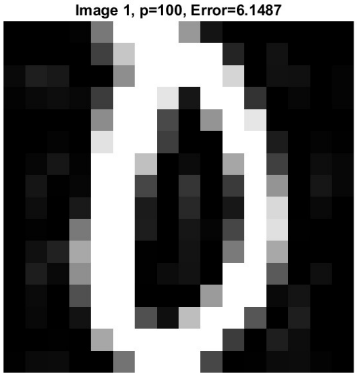


Image 1, p=100, Error=6.1487

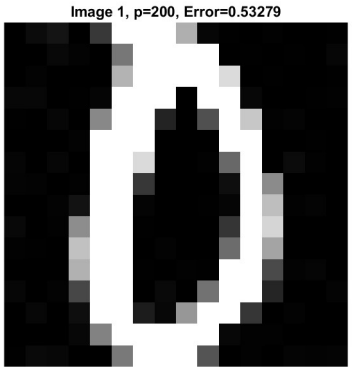


Image 1, p=200, Error=0.53279

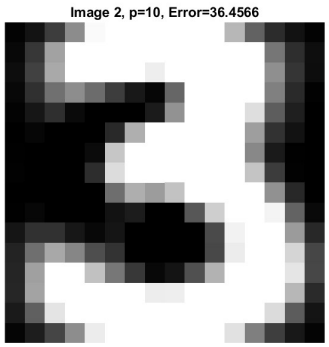


Image 2, p=10, Error=36.4566

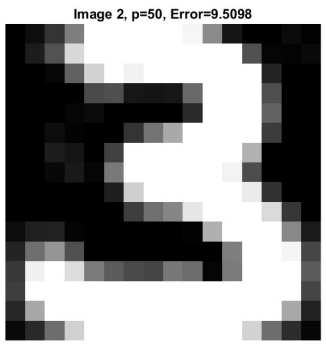


Image 2, p=50, Error=9.5098

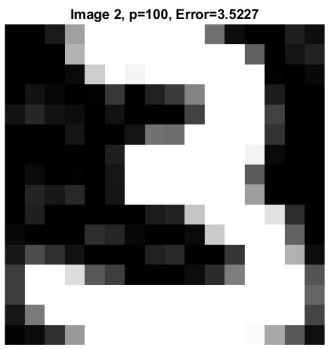


Image 2, p=100, Error=3.5227

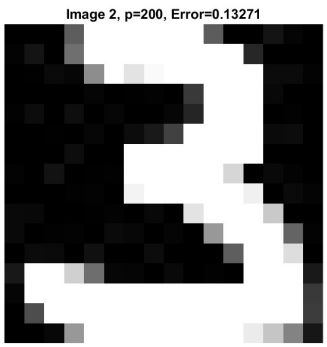


Image 2, p=200, Error=0.13271