

## CSCI567 2013 Homework Assignment 5

### Programming Report

Student Name	Last 4 digits of USC ID
Jia Li	0854
Udit Agrawal	5165
Yin-Ray Rick Huang	6794
Arpit Bansal	0979

# 1 Anomaly Detection

## 1.1 Kmeans-1

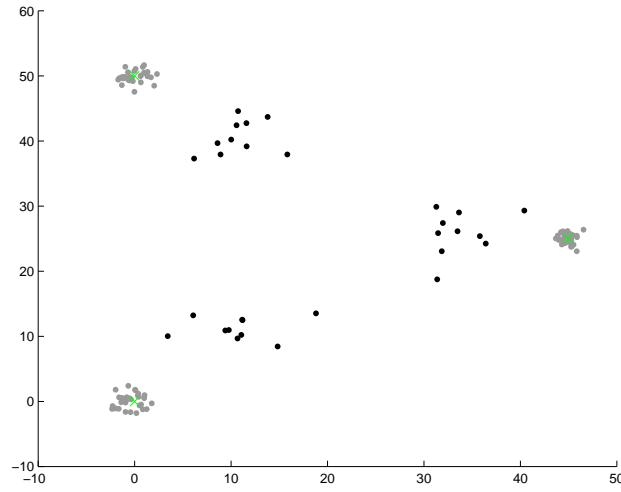


Figure 1: All the Data Points.  $Means = [[0, 0], [0, 50], [45, 25]]$

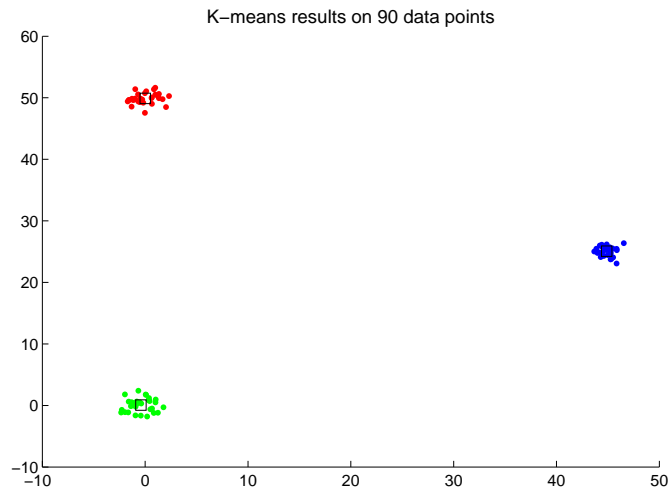


Figure 2:  $d_{red} = 0.0107$ ;  $d_{green} = 0.1739$ ;  $d_{blue} = 0.0187$

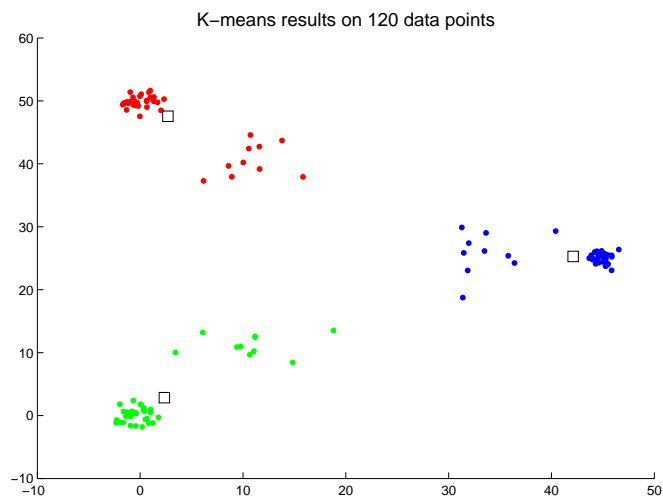


Figure 3:  $d_{red} = 13.2583$ ;  $d_{green} = 13.6640$ ;  $d_{blue} = 8.4678$

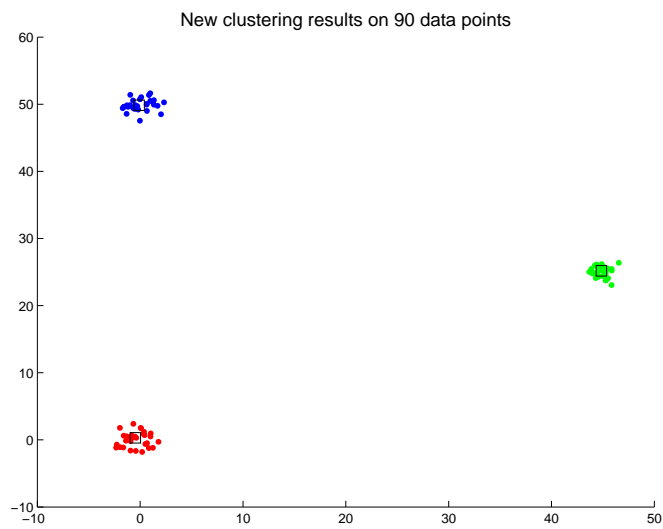


Figure 4:  $d_{red} = 0.0322$ ;  $d_{green} = 0.3129$ ;  $d_{blue} = 0.0549$

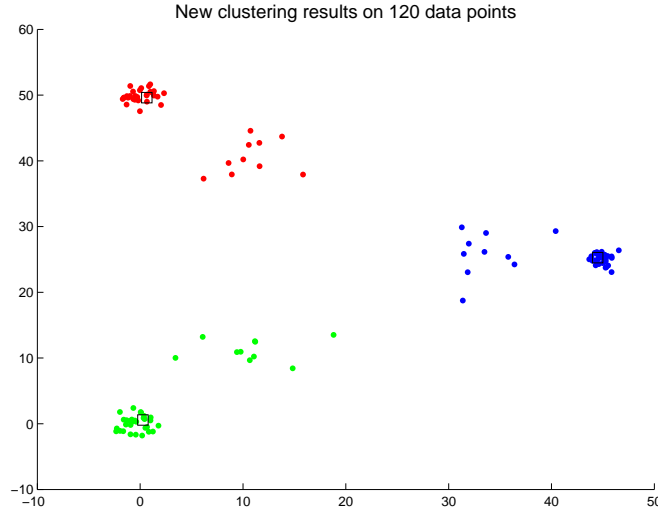


Figure 5:  $d_{red} = 0.5685$ ;  $d_{green} = 0.4041$ ;  $d_{blue} = 0.3328$

It can be seen from the figure 3 and figure 5, when we added the outlier points into the data, the mean change a log for the L2-Kmeans, but for L1-Kmeans. The distance of new means are still very close to old one. So L1-Kmeans is much more robust than the L2 one. The distance in plot 5 is only slightly larger than the plot 4, so the outlier have make too much difference for the center of cluster.

## 2 Kmeans on Photo Compression

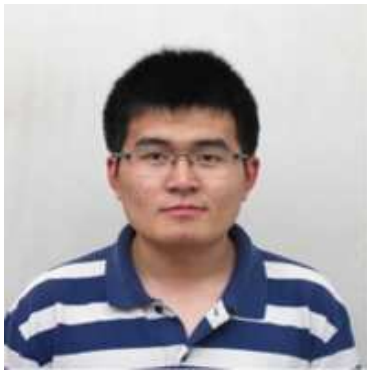


Figure 6: Original Picture

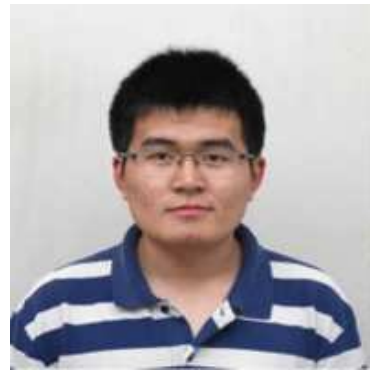


Figure 7: K=400

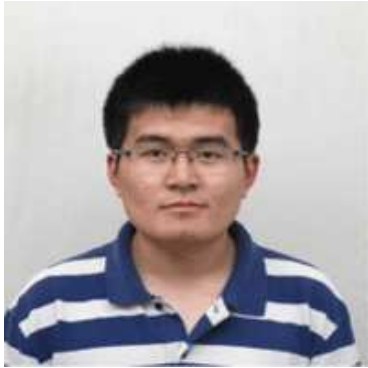


Figure 8:  $K=200$

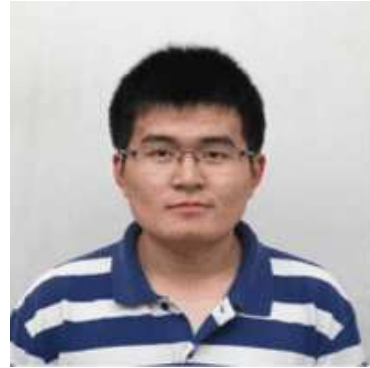


Figure 9:  $K=100$



Figure 10:  $K=10$

### 3 PCA for Preprocessing

#### 3.1 Diagonized Prediction

M	5	20	40	60	100
C	3.2	1.6	1.6	1.6	6.4
Gamma of RBF	4e-6	4e-6	2e-6	2.67e-6	8e-7
Test Accuracy	0.4885	0.9000	0.9310	0.9285	0.9325

Table 1: Optimal Parameters and Testing Accuracy

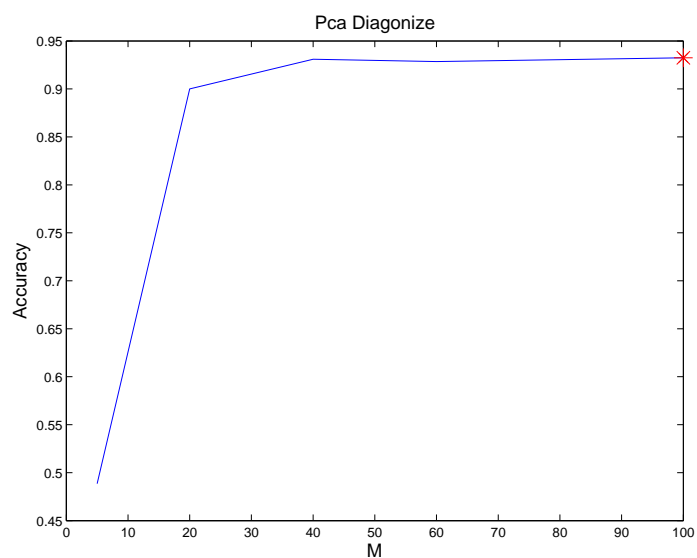


Figure 11: Best Accuracy = 0.9325,  $M^* = 100$

#### 3.2 Whiten Prediction

M	5	20	40	60	100
C	1.6	0.8	1.6	1.6	1.6
Gamma of RBF	0.32	0.08	0.04	0.0267	0.016
Testing Accuracy	0.4875	0.9275	0.9555	0.9535	0.9330

Table 2: Optimal Parameters and Testing Accuracy

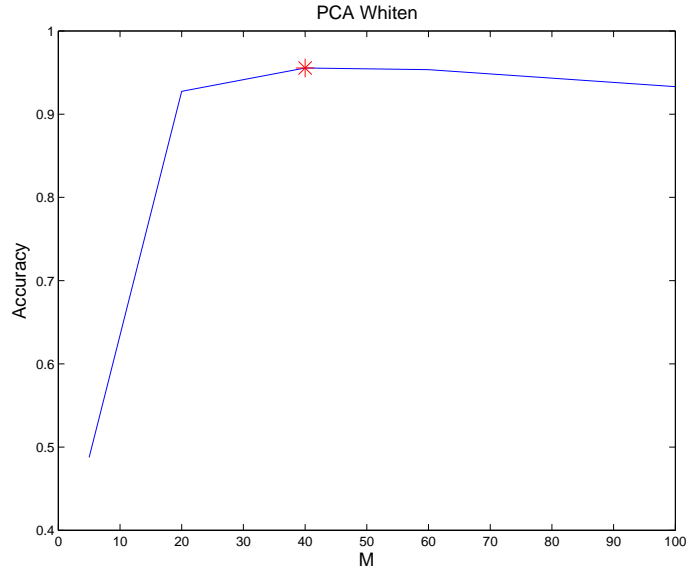


Figure 12: Best Accuracy = 0.9555,  $M^* = 40$

### 3.3 Standard Prediction

Optimal Parameter:

$C = 3.2$ ;  $\text{Gamma} = 0.0064$

Testing Accuracy = 0.9035

This method is not as good as PCA. The main reason is that when PCA reduce the feature size, it will highlight some important features, and ignore the irrelevant feature, kind of feature selection. But this standardization method, treat every feature with equal weight by make each feature have 0 mean, and 1 variance.