CSE 446: Homework 4

ayush29f@cs.washington.edu
Ayush Saraf

Problem 1.1

$$u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Problem 1.2

$$\hat{x}_1 = -\sqrt{2}$$
$$\hat{x}_2 = 0$$
$$\hat{x}_3 = \sqrt{2}$$
$$\mu = 0$$
$$\sigma^2 = \frac{(\sqrt{2}-0)^2 + 0 + (-\sqrt{2}-0)^2}{3} = 4/3$$

Problem 1.3

$$error = \sum_{i=1}^{3} (\mathbf{x_i} - \hat{x}_i \mathbf{u_i})^2$$
$$error = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Problem 2.M.1

$$r_{ic} = \frac{\pi_c N(x_i|\mu_c \sigma_c)}{\sum_{j=1}^{C} \pi_j N(x_i|\mu_c \Sigma_j)}$$

$$l(\Sigma, \mu, \pi|x) = \sum_{c=1}^{C} \prod_{i}^{N=3} r_{ic}$$

Problem 2.M.2

$$\pi_c = \frac{N_c^{soft}}{N}$$

$$\pi_1 = \frac{1.4}{3}; \pi_2 = \frac{1.6}{3}$$

Problem 2.M.3

$$\mu_c = \frac{1}{N_c^{soft}} \sum_{i=1}^{N} r_{ic} x_i$$

$$\mu_1 = \frac{1}{1.4} \sum_{i=1}^{3} r_{i2} x_i = \frac{1+4+0}{1.4} = \frac{5}{1.4} = 3.57$$

$$\mu_2 = \frac{1}{1.6} \sum_{i=1}^{3} r_{i2} x_i = \frac{0+6+20}{1.6} = \frac{26}{1.6} = 16.25$$

Problem 2.M.4

$$\Sigma_c = \frac{1}{N_c^{soft}} \sum_{i=1}^{N} r_{ic} (x_i - \hat{\mu_c})^2$$

$$\Sigma_1 = \frac{1}{1.4} \sum_{i=1}^{3} r_{i1} (x_i - \hat{\mu_1})^2$$

$$\Sigma_1 = \frac{(3.6/1.4)^2 + 0.4 * (9/1.4)^2}{1.4} = 16.53$$

$$\sigma_1 = 4.06$$

$$\Sigma_2 = \frac{1}{1.6}\sum_{i=1}^{3} r_{i1}(x_i - \hat{\mu_2})^2$$

$$\Sigma_2 = \frac{0.6 * (10/1.6)^2 + (6/1.6)^2}{1.6} = 23.44$$

$$\sigma_2 = 4.84$$

Problem 2.E.1

$$r_{ic} = \frac{\pi_c N(x_i|\mu_c \sigma_c)}{\sum_{j=1}^{C} \pi_j N(x_i|\mu_c \Sigma_j)}$$

Problem 2.E.2

$$r = \begin{bmatrix} 0.9919 & 0.0081 \\ 0.4075 & 0.5925 \\ 0.0004 & 0.9996 \end{bmatrix}$$

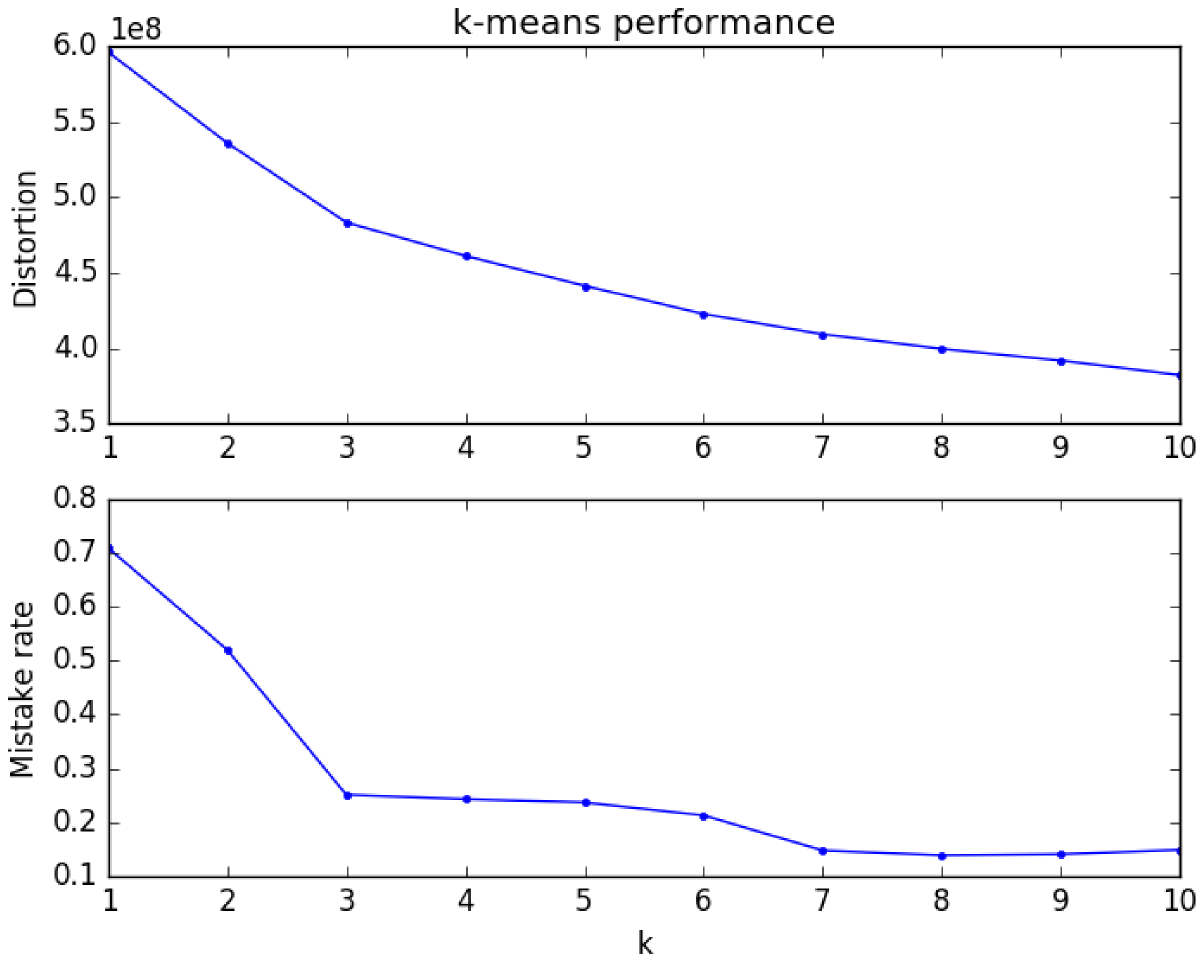
Problem 3.5

After looking at the performance plots, I investigated a little more by picking a set of 3 features with highest variance in the dataset to see how things are being classified. I plotted these features with different k values and also individual plots of ground truth of each label.

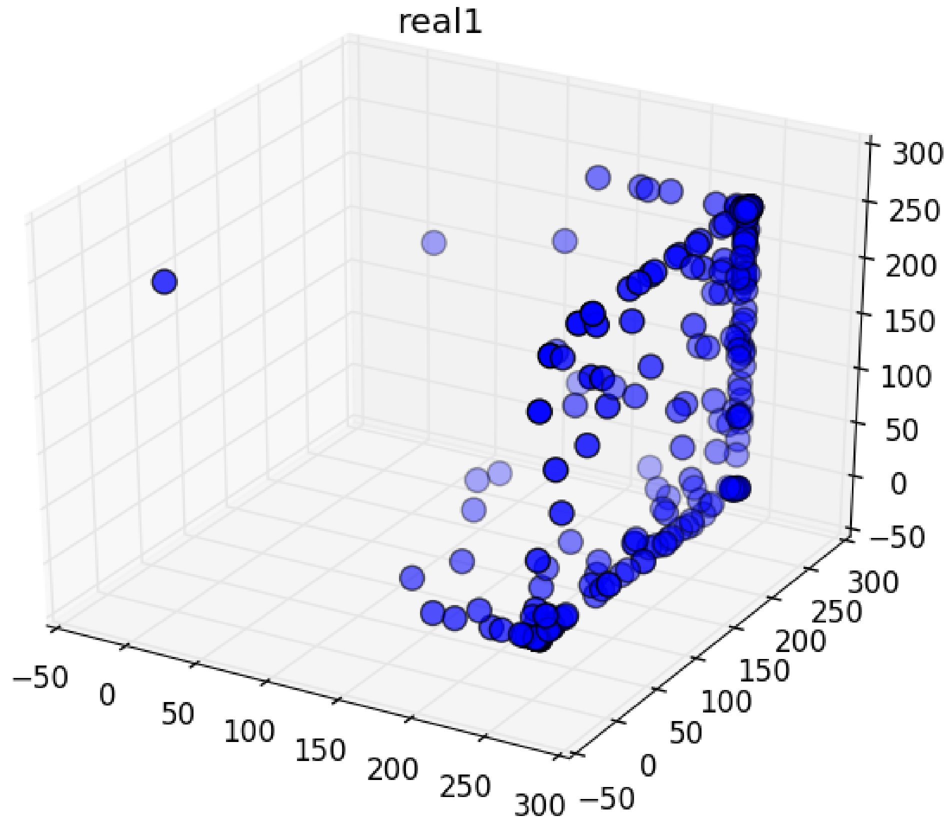
Looking at the graphs, it makes sense why the mistake rate decreased dramatically up until k=3. It can be seen that images with label 1, 7 are quite easily differentiable from rest even in these 3 dimensions. However, we can see it is still very hard to differentiate between 5 and 3 because they are closely together in this feature space. Which is why the accuracy didn't increase that much when we moved to k=4. However, there is some improvement when we hit k=7, this is essentially because of higher number of clusters with multiple clusters identifying same thing, we were able to get a better accuracy. Although this is very random and at different runs we find this wiggle in different k values.

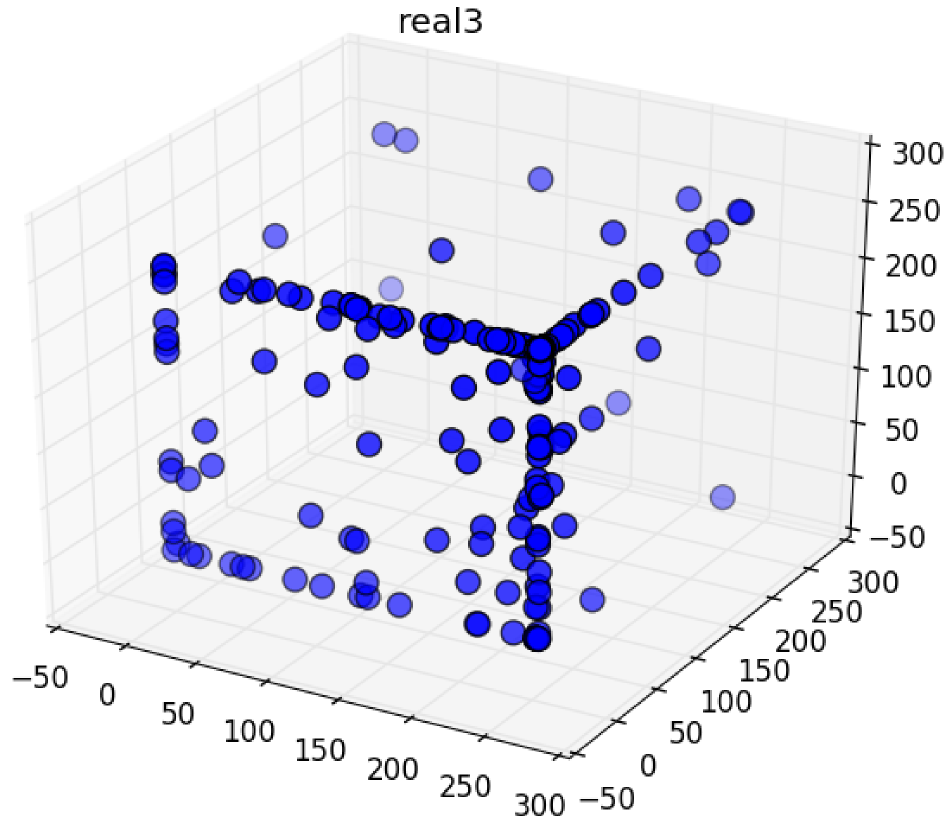
The with-in group sum of squared would decreases because there are more cluster centers and therefore its more likely for these points to be close the cluster centers.

All these conclusions matched with my intuition about the problem and the plots confirmed it.

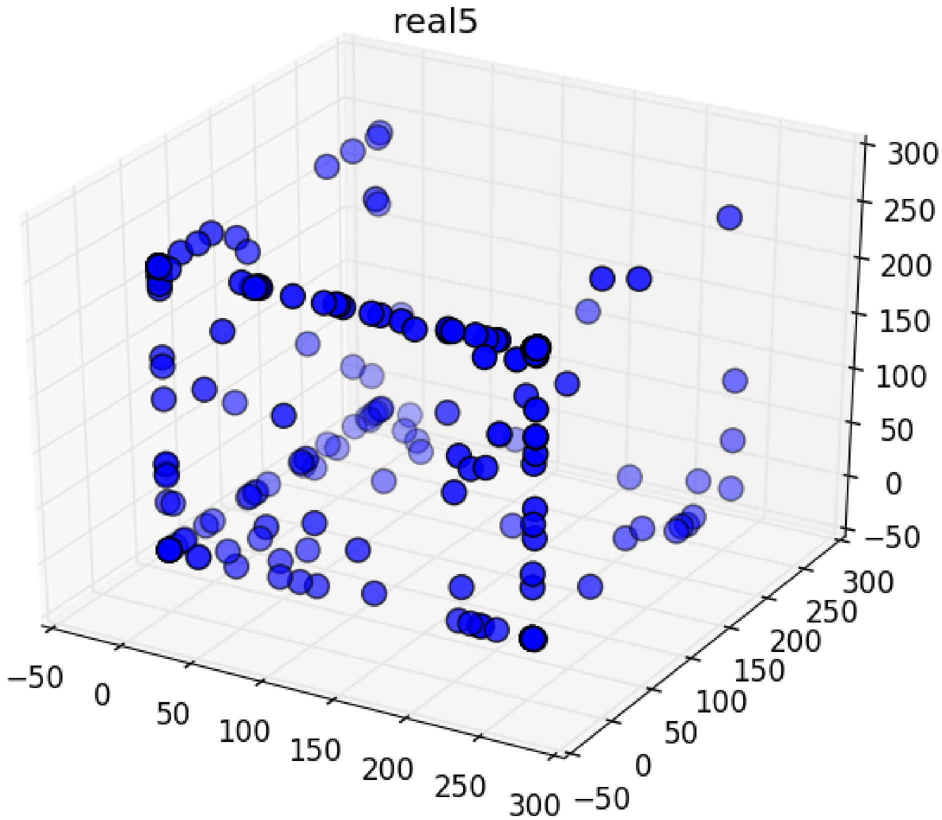


k-means performance
1e8
Distortion
6.0
5.5
5.0
4.5
4.0
3.5
Mistake rate
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1
2
3
4
5
6
7
8
9
10
k

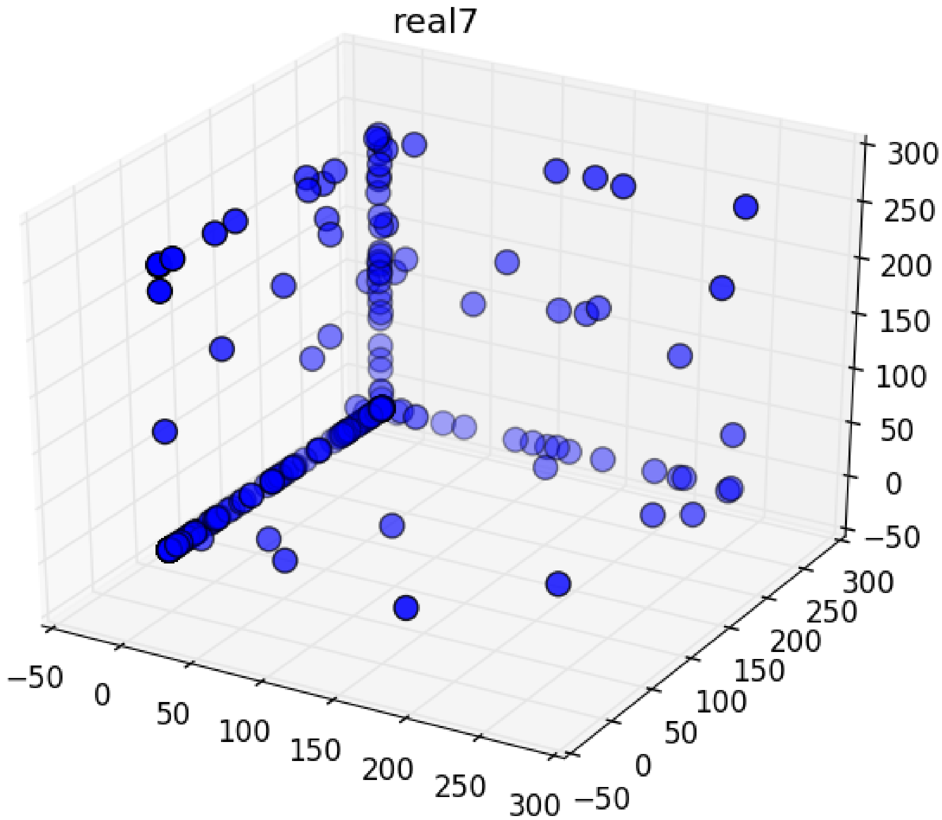

real1
300
250
200
150
100
50
0
−50
−50
0
50
100
150
200
250
300
300
250
200
150
100
50
0
−50

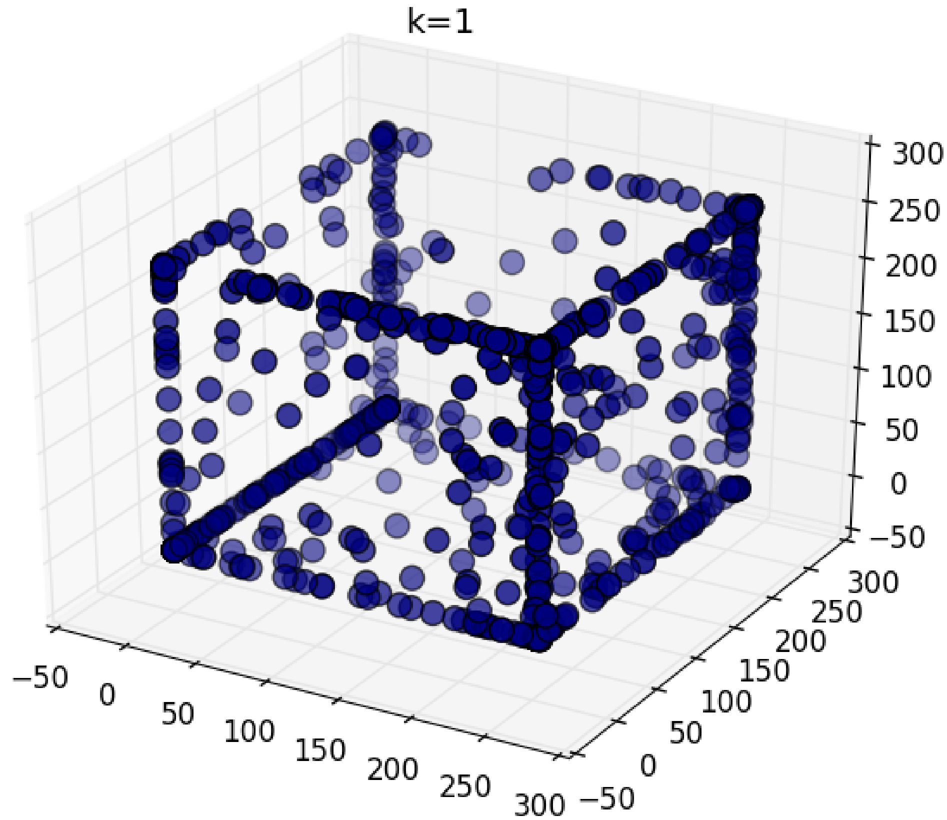


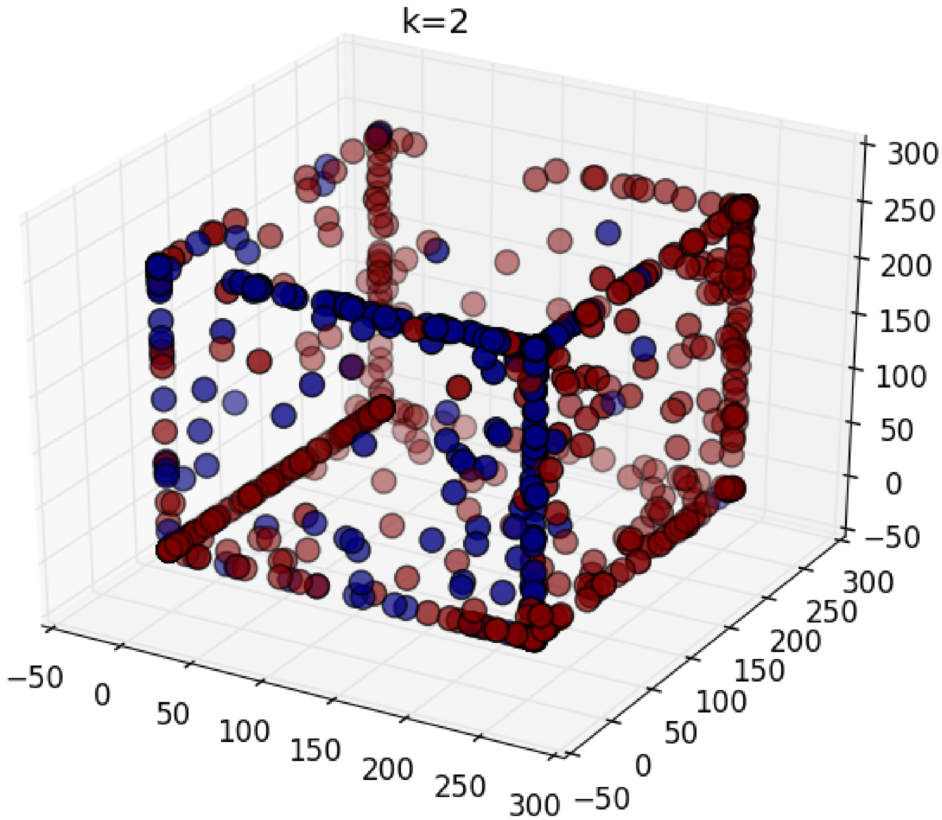
real3
300
250
200
150
100
50
0
−50
−50
0
50
100
150
200
250
300
300
250
200
150
100
50
0
−50

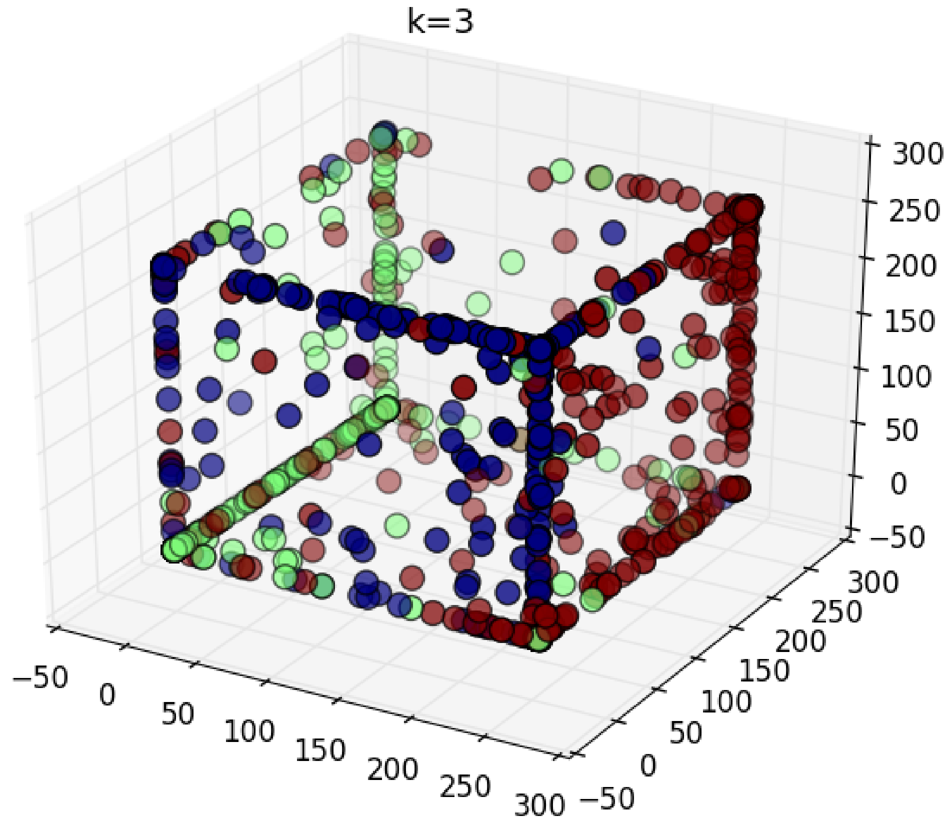


real5
−50
0
50
100
150
200
250
300

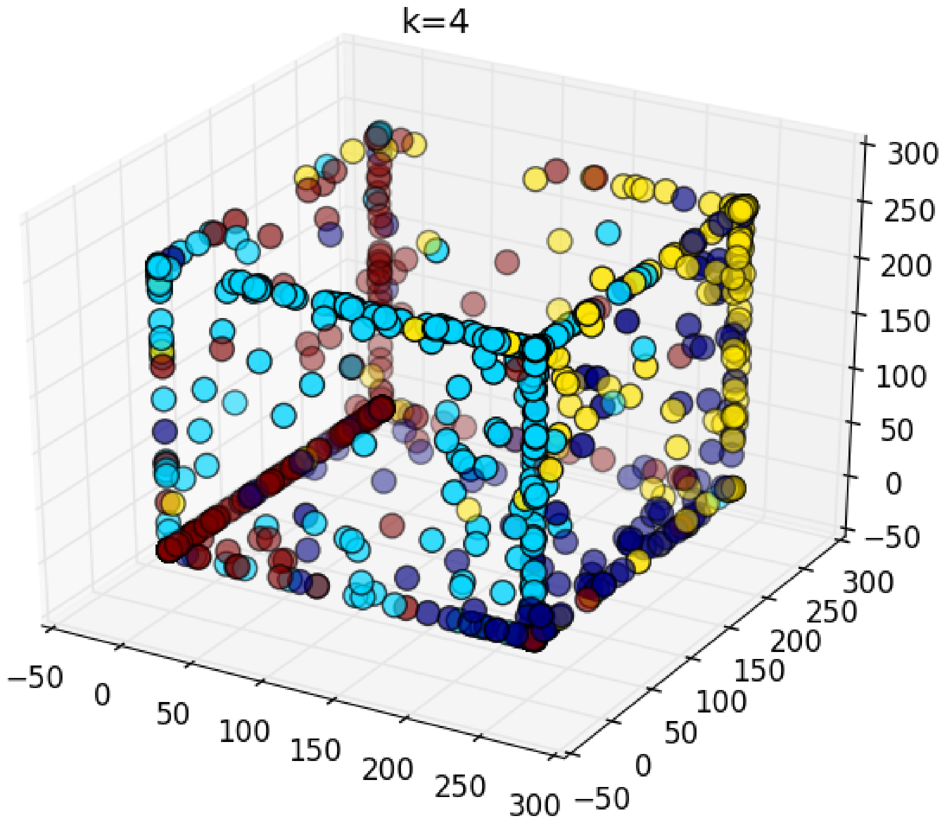

real7
−50
0
50
100
150
200
250
300
300
250
200
150
100
50
0
−50
300
250
200
150
100
50
0
−50

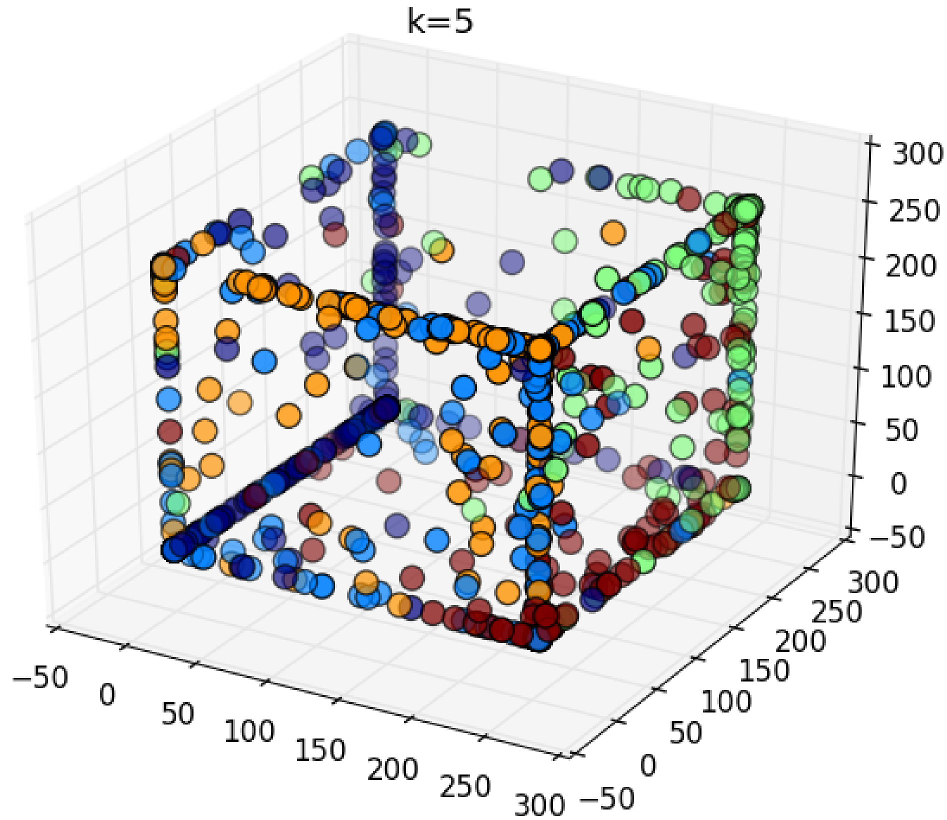

k=1
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300

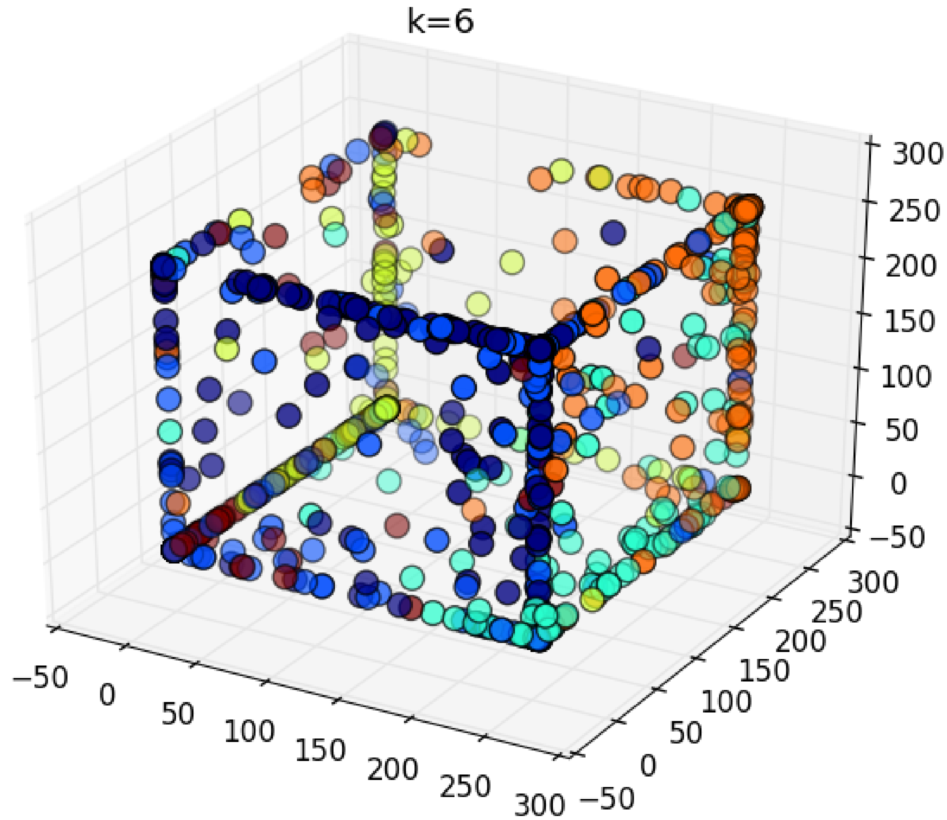

k=2
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300

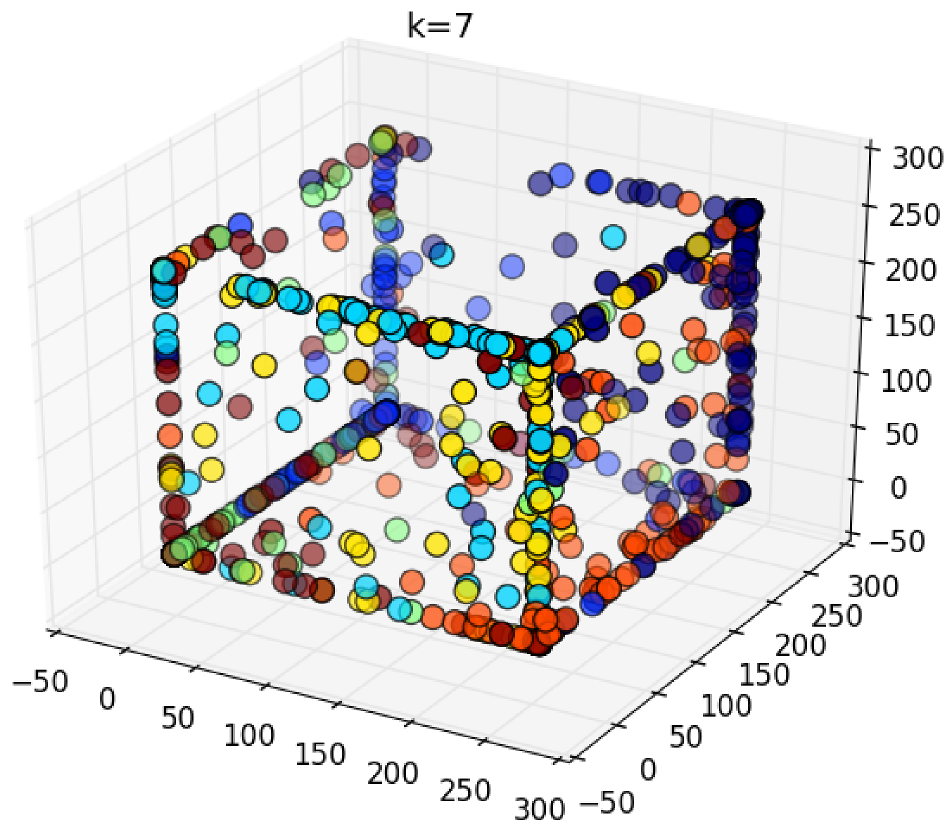


k=3
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300

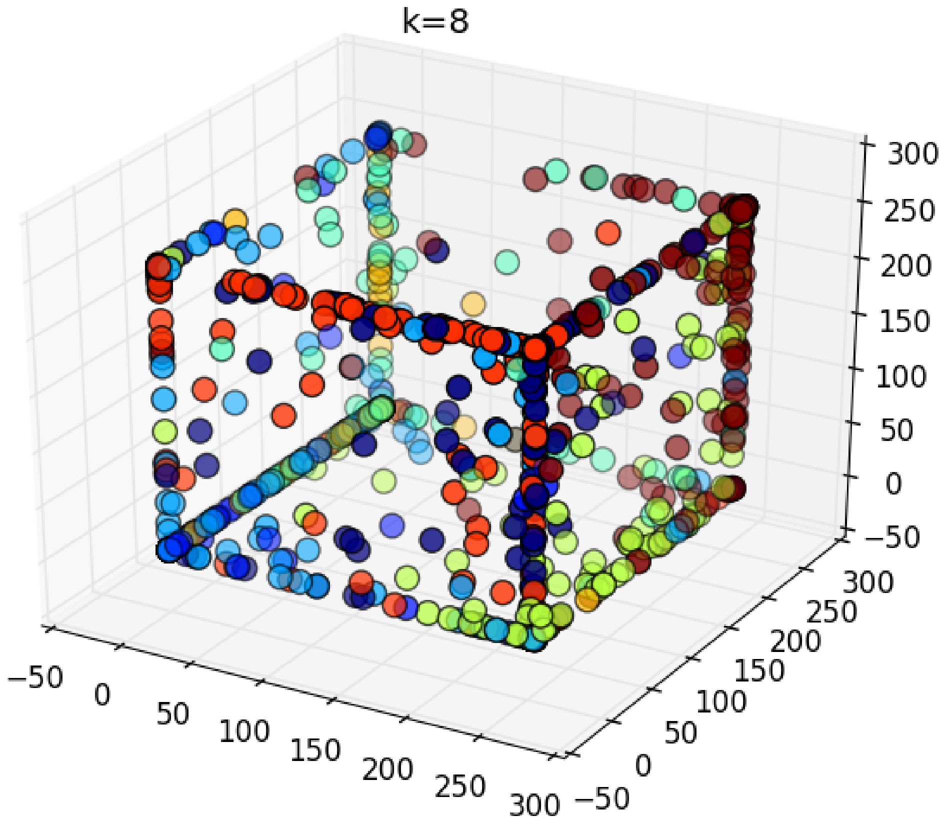

k=4
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300


k=5
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300

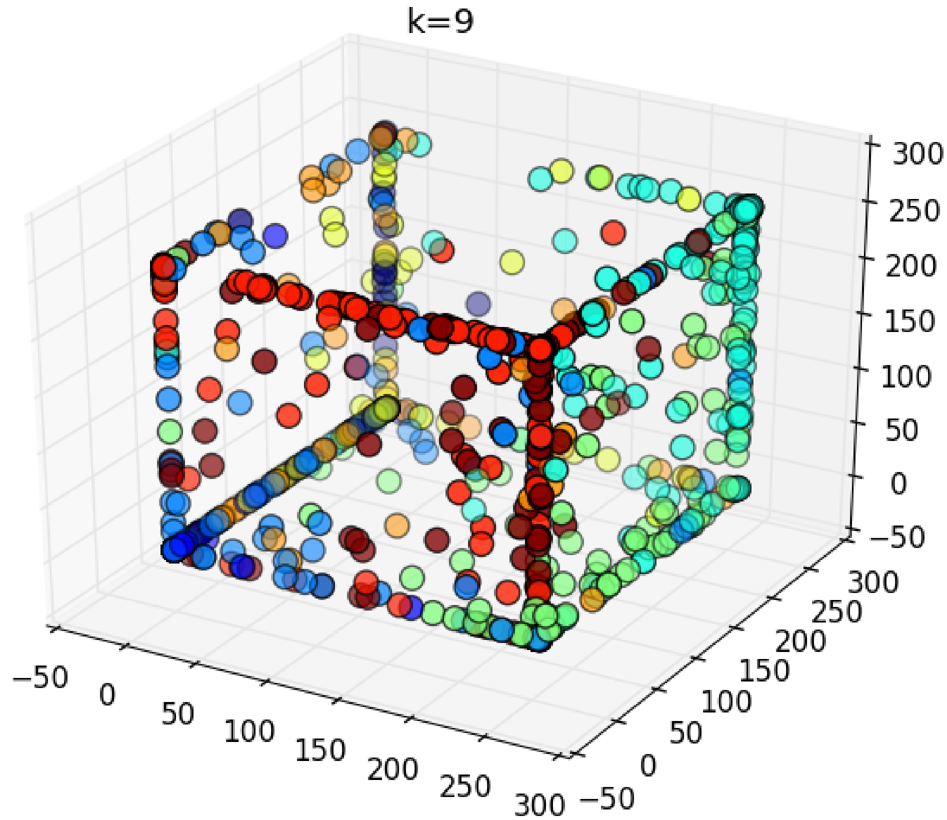

k=6
300
250
200
150
100
50
0
−50
−50
0
50
100
150
200
250
300
300
250
200
150
100
50
0
−50



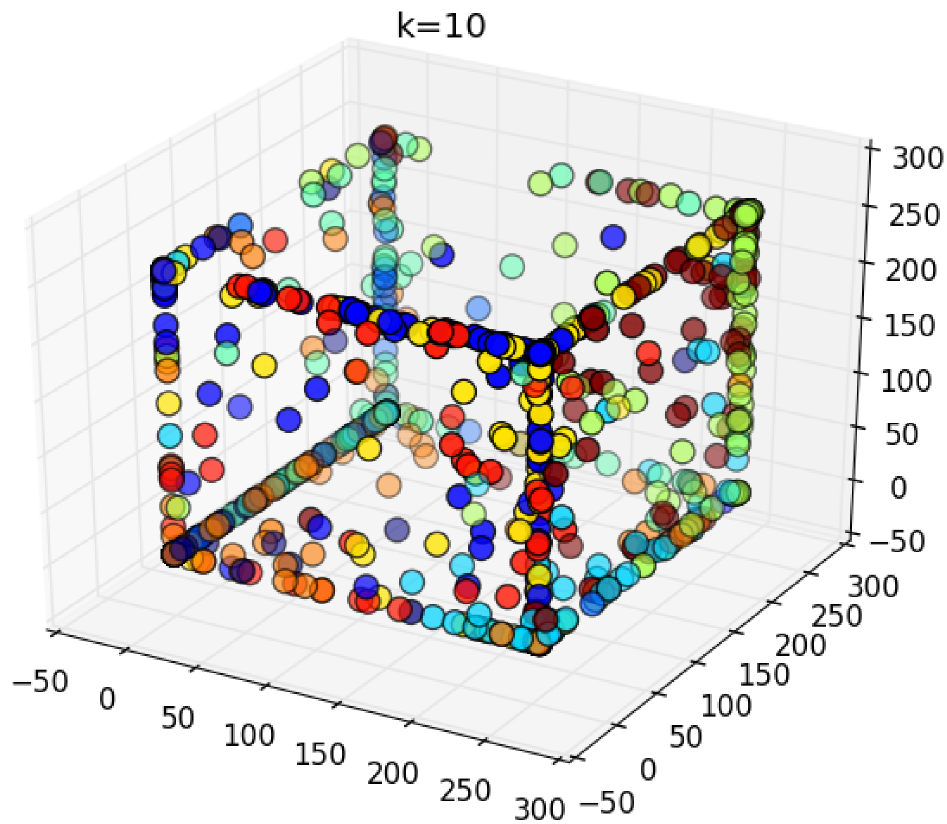
k=7



k=8
300
250
200
150
100
50
0
−50
300
250
200
150
100
50
0
−50
−50
0
50
100
150
200
250
300



k=9
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300
−50
0
50
100
150
200
250
300


k=10
−50
0
50
100
150
200
250
300