

Clusterings

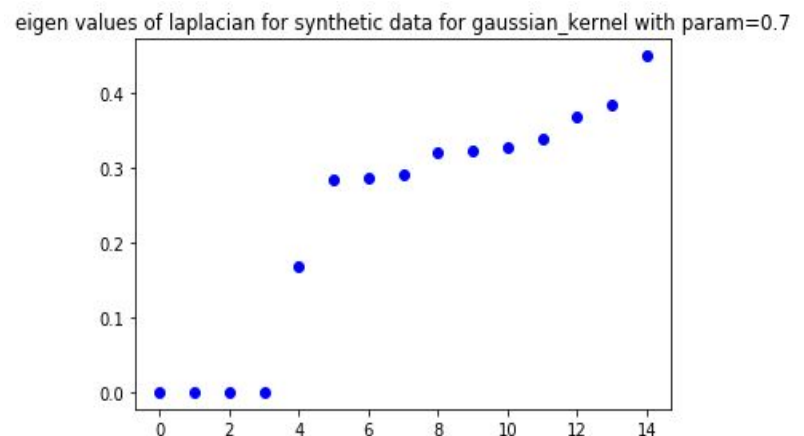
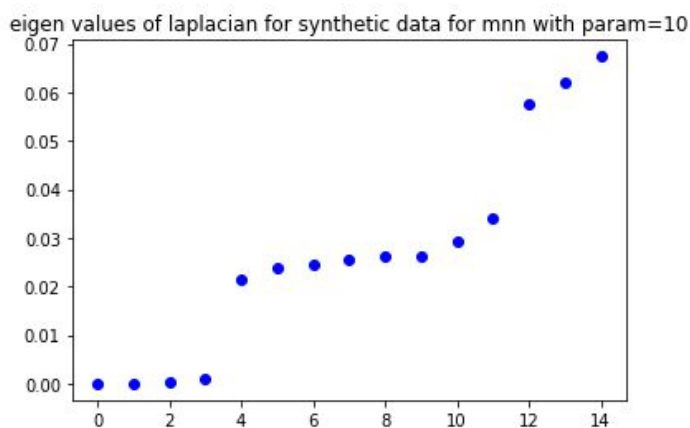
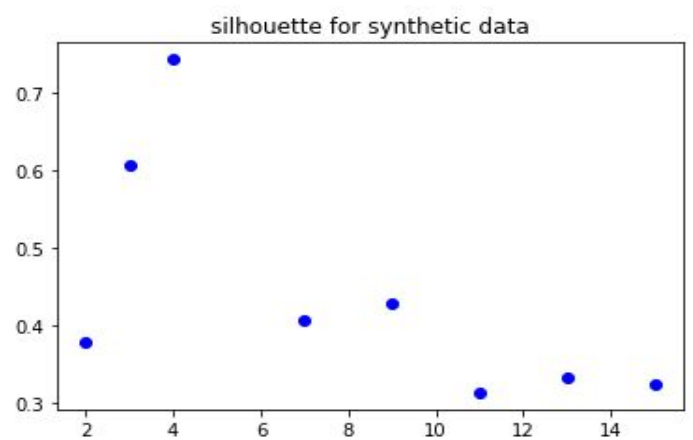
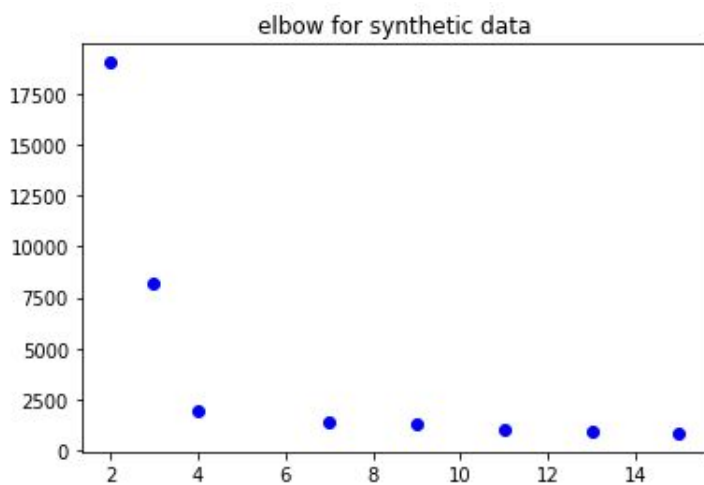
ID : 305248791

Synthetic dataset:

I created synthetic dataset with 4 clusters using make_blobs from sklearn.

K selection:

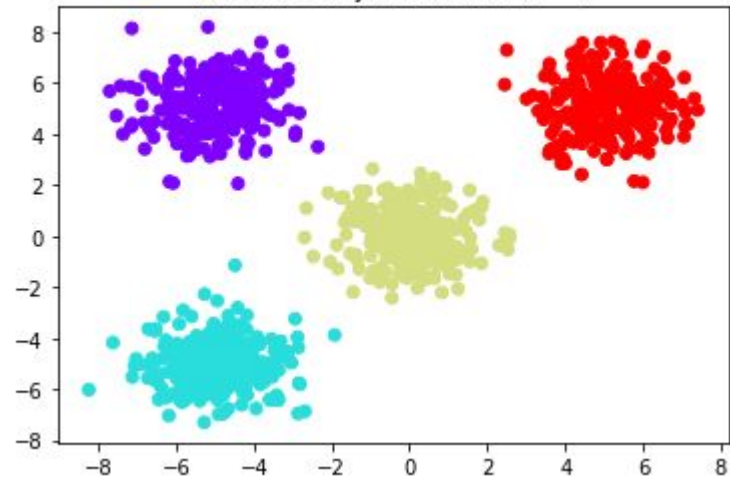
Results of applying different K selection methods (elbow, Silhouette, eigen gap).



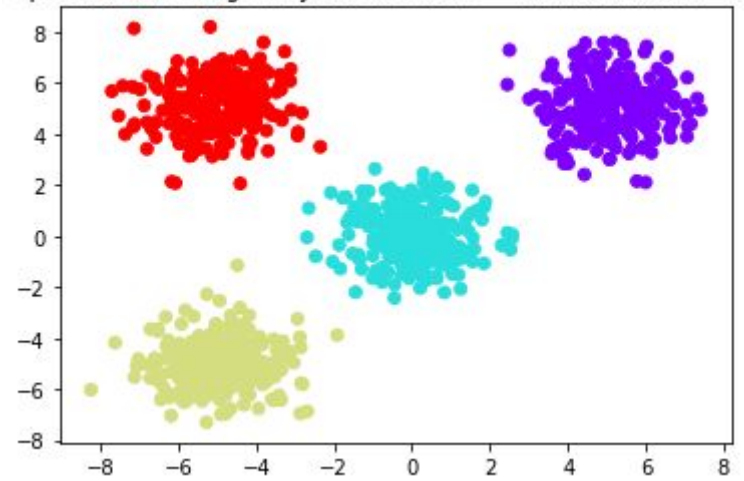
We can see that in the elbow method the decreasing start to plateau at $k = 4$, and the maximum silhouette is at $k=4$, and the first gap at eigen gap for both similarities is at $k = 4$, so all methods agree on $k=4$ as the k to choose and the data itself is composed of 4 clusters which makes sense.

Kmeans and spectral clustering:

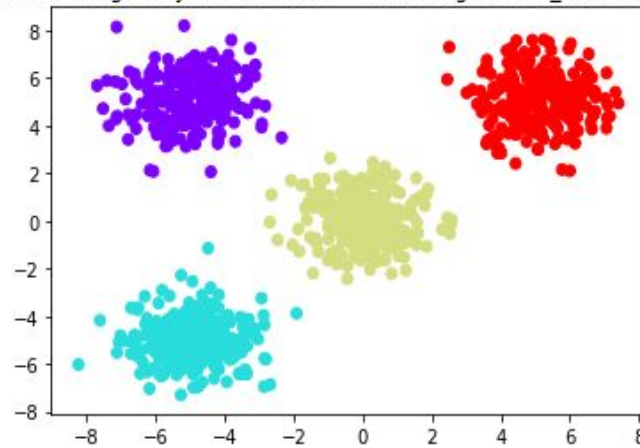
kmeans on synthetic data $k = 4$



spectral clustering on synthetic data $k = 4$, with mnn, $m = 10$



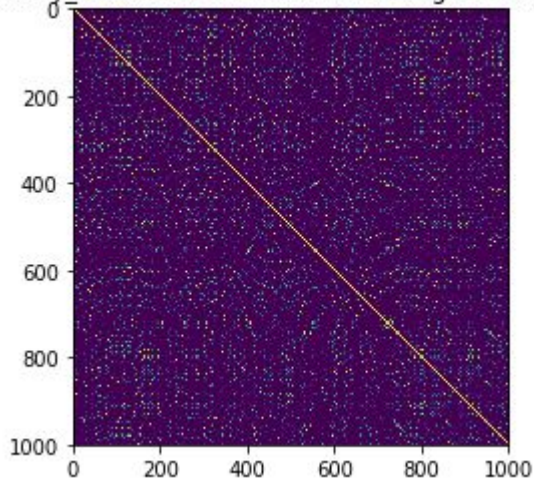
spectral clustering on synthetic data $k = 4$, with gaussian_kernel, $\sigma = 0.7$



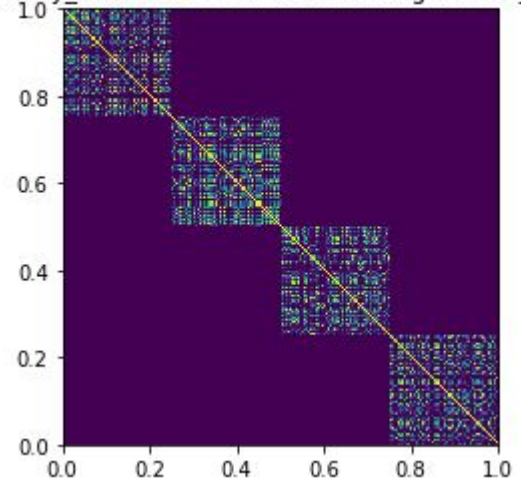
The data was successfully clustered by kmeans and spectral clustering on both similarities.

Plotting the Similarity Graph:

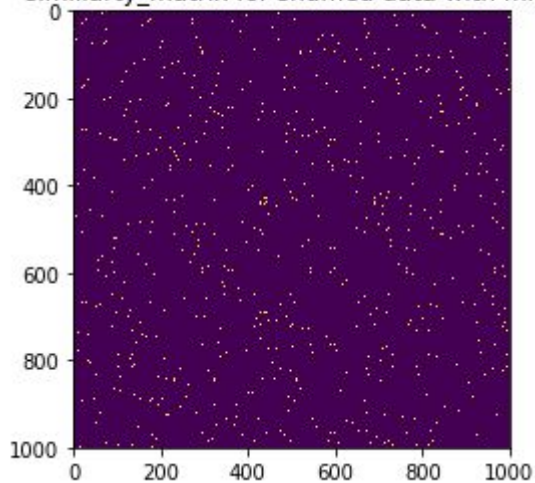
similarity_matrix for shuffled data with gaussian_kernel



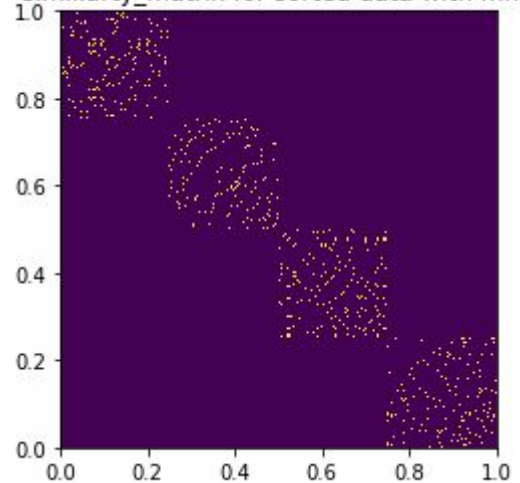
similarity_matrix for sorted data with gaussian_kernel



similarity_matrix for shuffled data with mnn

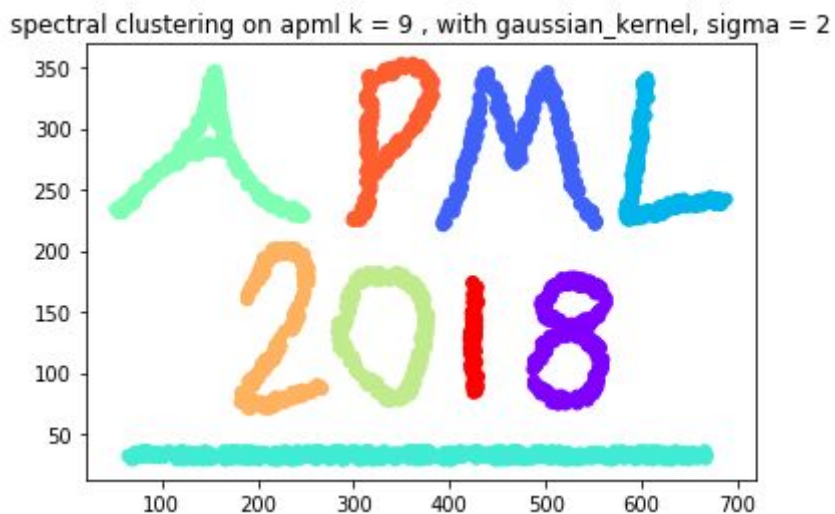
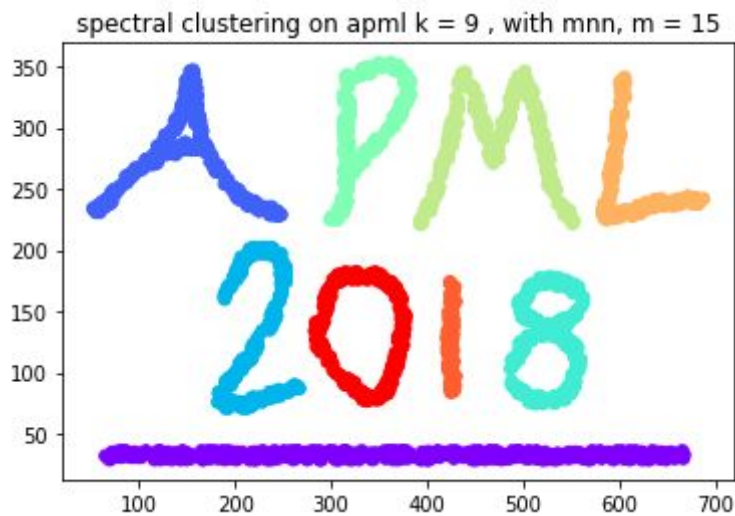
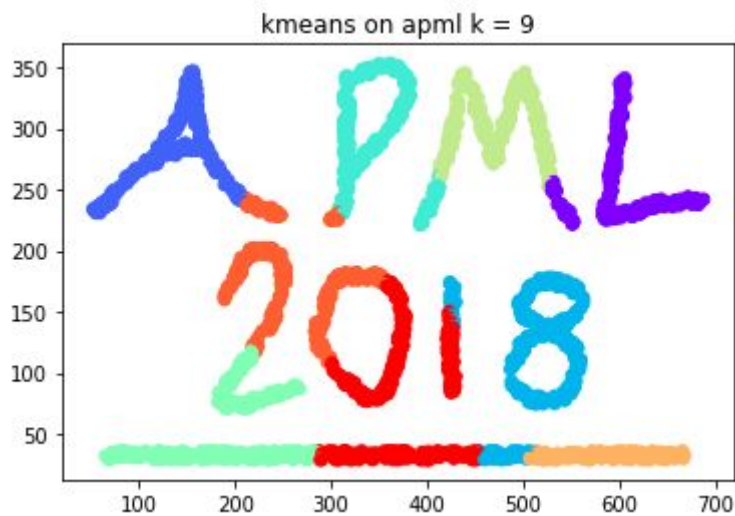


similarity_matrix for sorted data with mnn



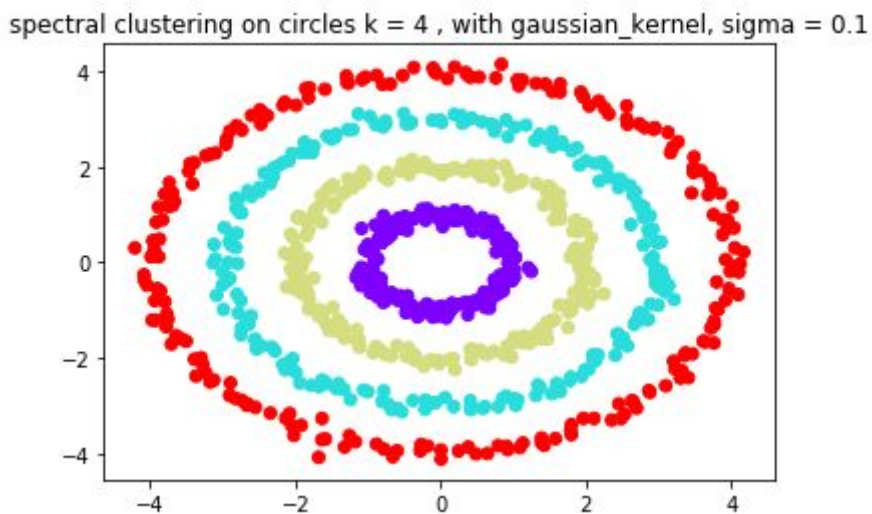
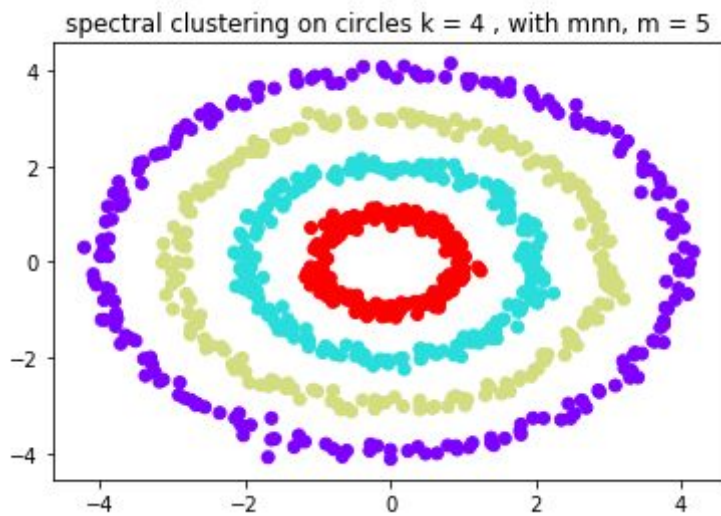
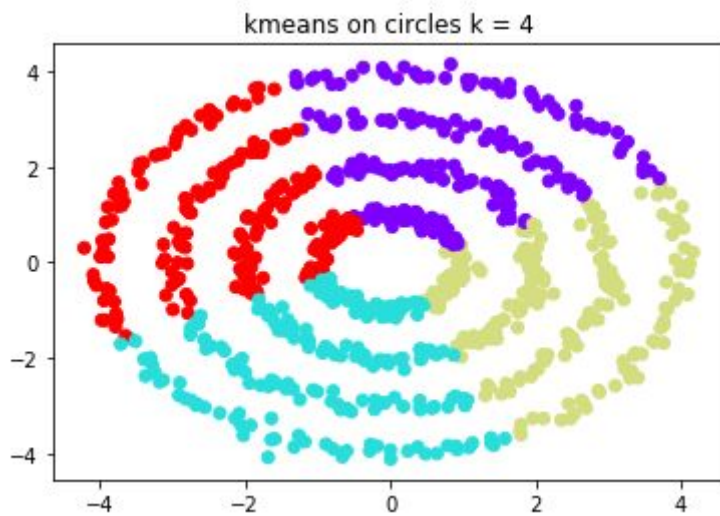
In right images we can see the 4 clusters represented by the 4 squares.
So the similarity graph is a good representation of the clustering.

APML dataset:



We can see that the spectral clustering on both similarities perform better than the K-means algorithm.

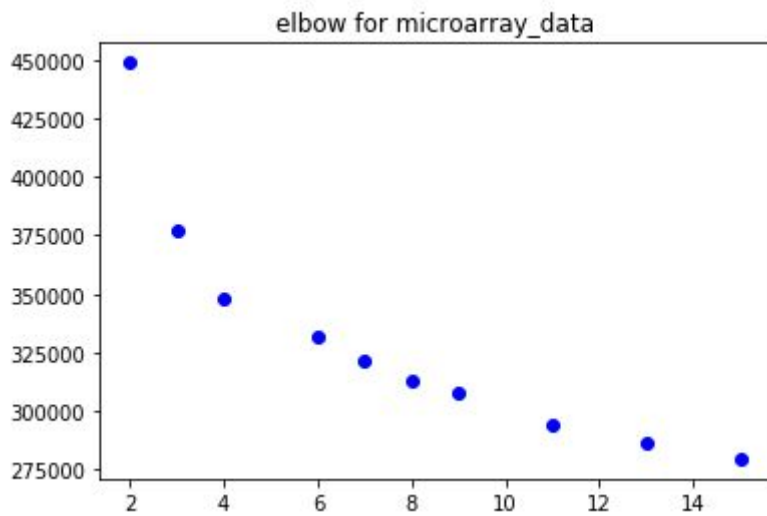
Circles dataset:



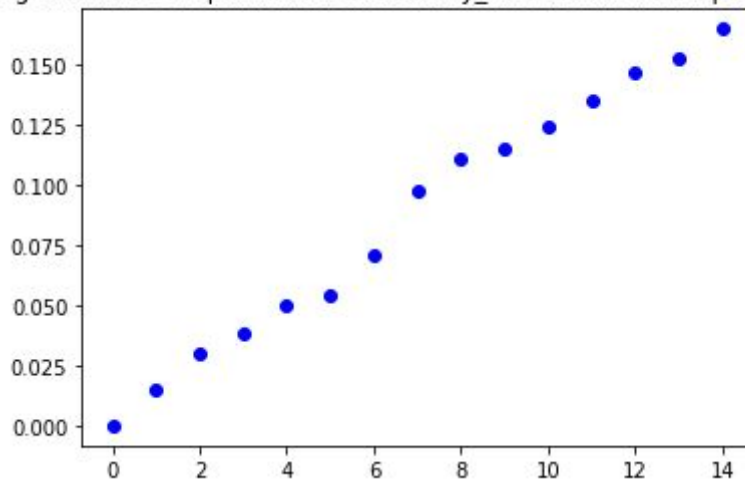
Here too we can see that the spectral clustering on both similarities perform better than the K-means algorithm.

Microarray dataset:

K selection:



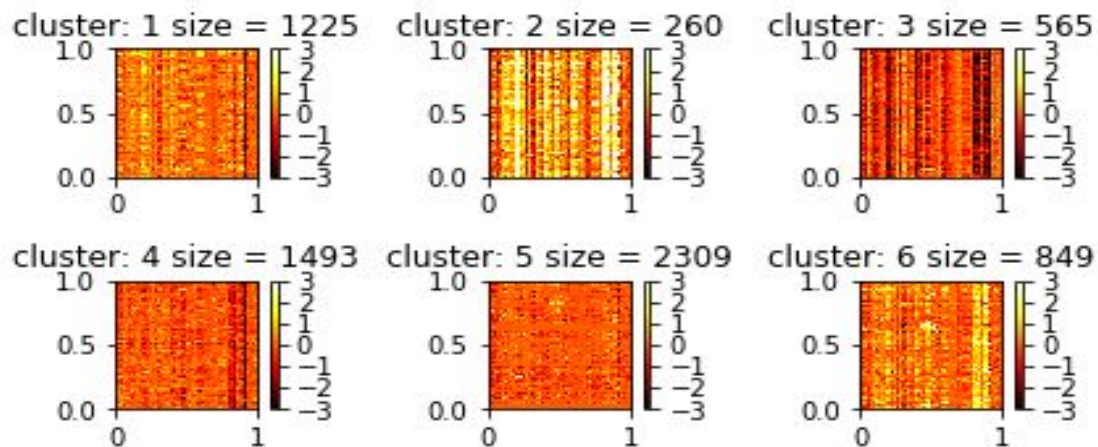
eigen values of laplacian for microarray_data for mnn with param=10



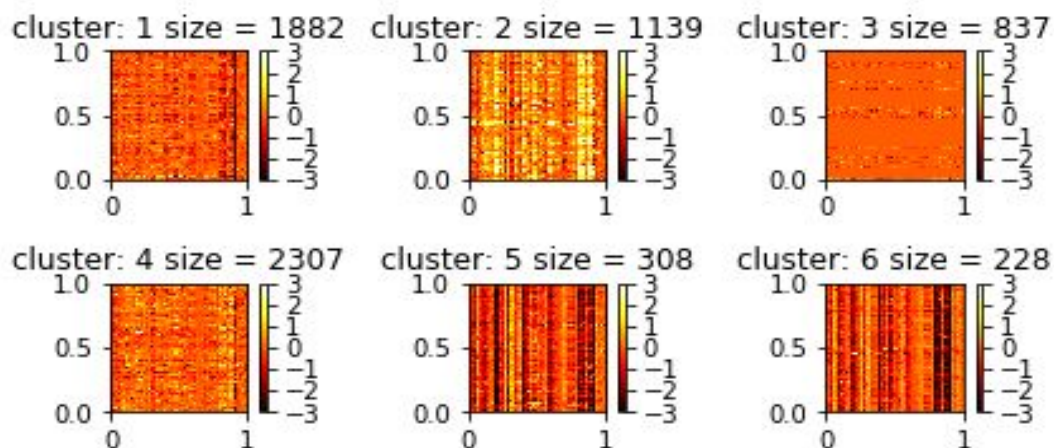
In elbow method at $k=6$ the graph starts to plateau, at the second graph the first noticeable eigen gap is at $k=6$. so I choose $k=6$ for the number of clusters.

$k=7$ and $k=8$ are good candidates too but I decided to stick with $k=6$ which gave good results as follows.

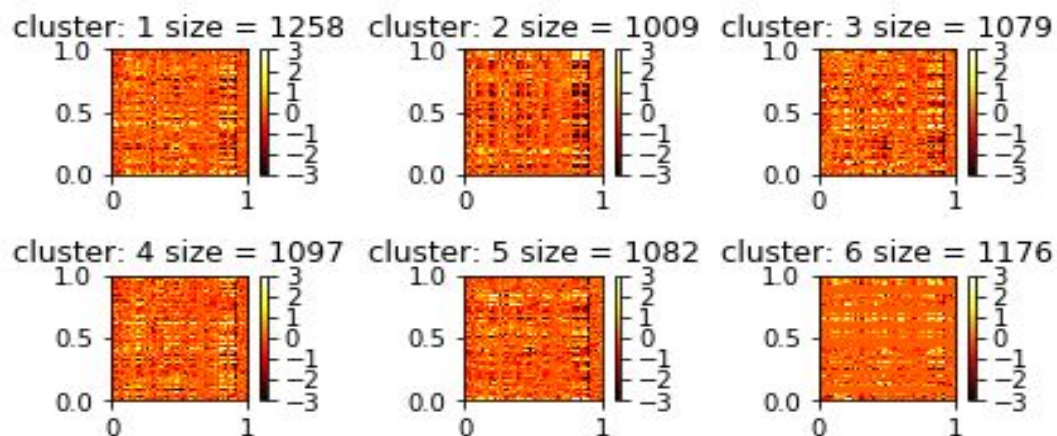
Kmeans for microarray data:



Spectral clustering with mnn with m=10:



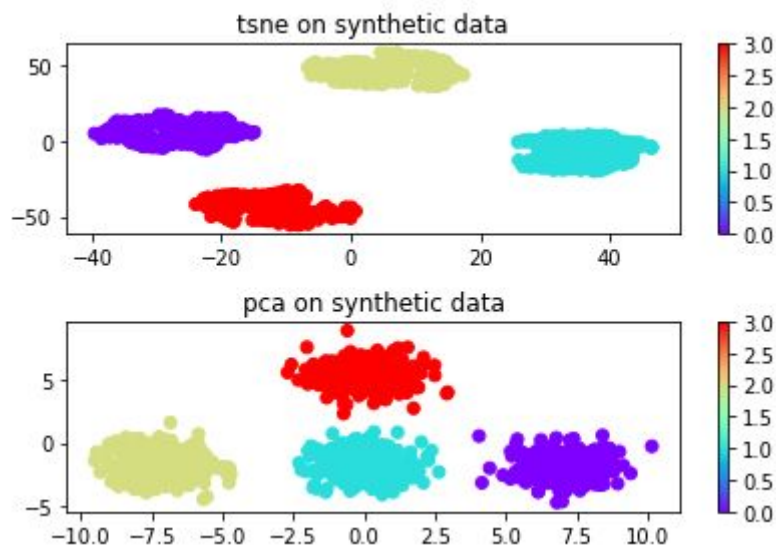
Spectral clustering with gaussian kernel with sigma =1.5:



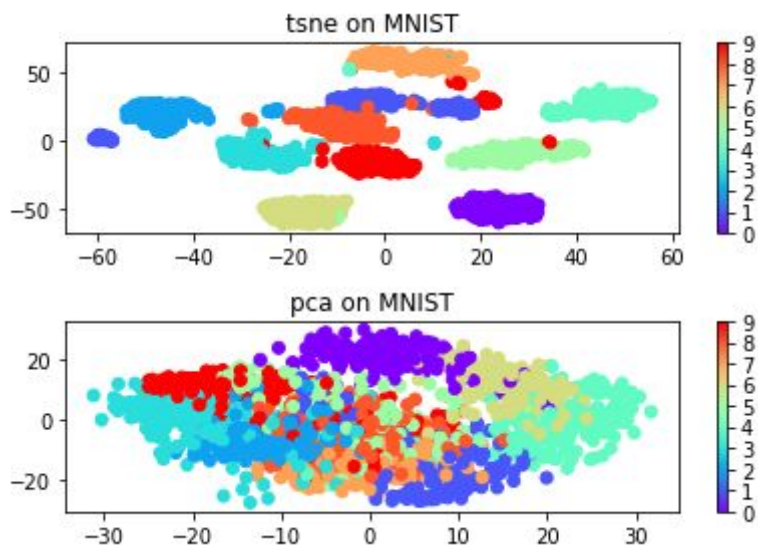
Best results is with mnn where $m=10$ we can the see clearest that the rows are similar especially at clusters 3,5,6.

t-SNE:

I created synthetic data with 4 cluster with 10 dimensions.



t-SNE clusters the synthetic data better.



On the MNIST data it is even more clear that the t-SNE clusters the digits better.

שאלה 7.7

נניח U היא סדרת מספרים סופית של וקטורים אנטי-אורתוגונליים

יכול להיות בהם, מספר סופי של וקטורים אנטי-אורתוגונליים

למספר האנטי-אורתוגונליים בין n ל- $n+1$ חזקת n לא

וקטורים אנטי-אורתוגונליים. נראה שמוקדית הווקטור $\sum_{i=1}^n \|x_i - u\|^2$, היא מינימום

יורד ב-1 ככל צעד של האלגוריתם:

1. בחינה - Centroid, \bar{x} - ממוצע וקטורים אנטי-אורתוגונליים

נבחרים כפי שמוקדית מינימום של וקטורים אנטי-אורתוגונליים, $\sum_{i=1}^n \|x_i - u\|^2$. נראה שמוקדית הווקטור יכולה להיות צמודה

2. בחינה - וקטורים אנטי-אורתוגונליים, כפי שמוקדית מינימום של וקטורים אנטי-אורתוגונליים

יורד ב-1 כפי שמוקדית הווקטור יכולה להיות צמודה. $C(x) = \arg \min_k (\|x - u_k\|^2)$, ציטוט

ללא שינוי מספר סופי של וקטורים אנטי-אורתוגונליים, ומוקדית

יורד ב-1 כפי שמוקדית הווקטור יכולה להיות צמודה. וקטורים אנטי-אורתוגונליים

בעניין. ציטוט שמוקדית הווקטור יכולה להיות צמודה. וקטורים אנטי-אורתוגונליים

וקטורים אנטי-אורתוגונליים, ציטוט שמוקדית הווקטור יכולה להיות צמודה.

①

1.2 : 1/10

$\{(0,0), (0,1), (2,0), (3,1)\}$: $\{P, M, N, Y, Z, \dots\}$

$\{ (1,0), (1,1) \}$

י"ו. יבגיון יג-א/סט"מאי. כ"ש יבגיון באיבט"מ
י"ז. $\{(0,0.5), (2,0.5)\}$

? כמה ~~ה~~ 4 סטיות גורמות :
 $\{(0,0), (0,1), (1,0), (1,1)\}$
 ד"ר ע"י קולומביה אוסטרליה:

ד'תשנ"ח י"ג אלול ה'תשנ"ח

$$C_1 = \{(0,0), (0,1)\}, \quad C_2 = \{(1,0), (1,1)\}$$

p21

$$C_1 = \{(0,0), (1,0)\}, \quad C_2 = \{(0,1), (1,1)\}$$

②

1.3 > 1.0

1. $P = k! \prod \alpha_i$ $\Rightarrow E[\text{# trials}] = \frac{1}{k! \prod \alpha_i}$

$$P = k! \prod \alpha_i \Rightarrow E[\text{# trials}] = \frac{1}{k! \prod \alpha_i}$$

2.

$$E[\text{# trials}] = \frac{1}{k! \prod \alpha_i} \geq \min_{\alpha_i} \frac{1}{k! \alpha_i} = \frac{1}{k! \max_{\alpha_i} \prod \alpha_i}$$

$\prod \alpha_i$ \Rightarrow $\sum \alpha_i = 1$ \Rightarrow $\alpha_i = \frac{1}{k}$ \Rightarrow $\prod \alpha_i = \left(\frac{1}{k}\right)^k$

$$\arg \max_{\alpha_i} \prod \alpha_i = \arg \max_{\alpha_i} \log(\prod \alpha_i) = \arg \max_{\alpha_i} \left(\sum_{i=1}^k \log(\alpha_i) \right)$$

$$L(\alpha, \lambda) = \sum_{i=1}^k \log(\alpha_i) - \lambda \left(\sum_{i=1}^k \alpha_i - 1 \right)$$

λ \Rightarrow $\frac{\partial L}{\partial \alpha_i} = \frac{1}{\alpha_i} - \lambda = 0 \Rightarrow \alpha_i = \frac{1}{\lambda}$

$$\frac{\partial L}{\partial \alpha_i} = \frac{1}{\alpha_i} - \lambda = 0 \Rightarrow \alpha_i = \frac{1}{\lambda}$$

$$\sum_{i=1}^k \alpha_i = 1 \Rightarrow \sum_{i=1}^k \frac{1}{\lambda} = 1 \Rightarrow \frac{k}{\lambda} = 1 \Rightarrow \lambda = k \Rightarrow \alpha_i = \frac{1}{k} \quad \forall i$$

$$\Rightarrow \frac{1}{\lambda} \cdot k = 1 \Rightarrow \lambda = k \Rightarrow \alpha_i = \frac{1}{k} \quad \forall i$$

$$E[\text{# trials}] \geq \frac{1}{k! \max_{\alpha_i} \prod \alpha_i} = \frac{1}{k! \left(\frac{1}{k}\right)^k} = \frac{k^k}{k!}$$

(3)

הוא זה שיש לו את המידע, הוא זה שיש לו את המידע
הוא זה שיש לו את המידע, הוא זה שיש לו את המידע
הוא זה שיש לו את המידע, הוא זה שיש לו את המידע