Problem 1 (20 Points): Lloyd's Method

Given a dataset with seven data points $\{x_1, \cdots, x_7\}$ and the distances between all pairs of data points are in the following table.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0	5	3	1	6	2	3
x_2	5	0	4	6	1	7	8
x_3	3	4	0	4	3	5	6
x_4	1	6	4	0	7	1	2
x_5	6	1	3	7	0	8	9
x_6	2	7	5	1	8	0	1
x_7	3	8	6	2	9	1	0

Assume the number of clusters $k = 2$, and the cluster centers are initialized to be x_3 and x_6.

1. **5 Points.** What's the two clusters formed at the end of the first iteration of Lloyd's algorithm?
2. **5 Points.** What's the two clusters formed at the end of the second iteration of Lloyd's algorithm?
3. **10 Points.** What's the two clusters formed when the Lloyd's algorithm converges?

Problem 2 (15 Points): Guassian Mixture Model (GMM): Latent Variable View

Consider a GMM in which the marginal distribution $p(\mathbf{z})$ for the latent variable $\mathbf{z}$ is given by

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k},$$

where $\sum_{k=1}^{K} \pi_k = 1$; $\mathbf{z} = [z_1, z_2, \cdots, z_K]$ and z_k satisfies $z_k \in \{0, 1\}$ and $\sum_{k=1}^{K} z_k = 1$.
Moreover, the conditional distribution $p(\mathbf{x}|\mathbf{z})$ for the observed variable $\mathbf{x}$ is given by:

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)^{z_k}.$$

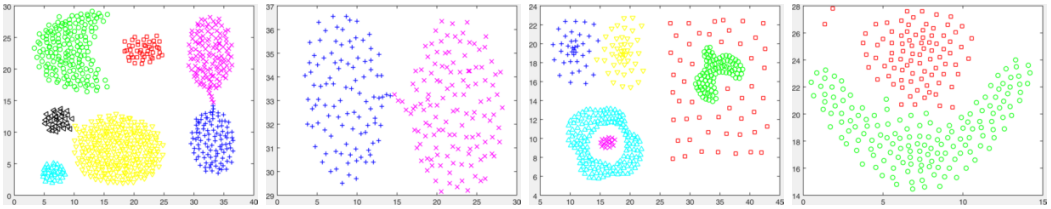
Prove that $p(\mathbf{x})$, obtaining by summing $p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$ over all possible values of $\mathbf{z}$, is a GMM. That is,

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z})p(\mathbf{x}|\mathbf{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k).$$

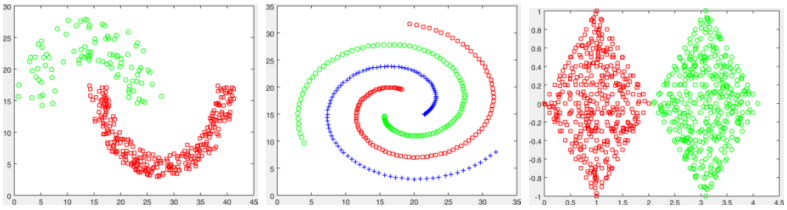
Problem 4 (20 Points): Generating GMMs

In this problem, you will write code to generate a mixture of 3 Gaussians satisfying the following requirements, respectively. Please specify the mean vector and covariance matrix of each Gaussian in your answer.

1. **6 Points.** Draw a data set where a mixture of 3 spherical Gaussians (where the covariance matrix is the identity matrix times some positive scalar) can model the data well, but K-means cannot.
2. **6 Points.** Draw a data set where a mixture of 3 diagonal Gaussians (where the covariance matrix can have non-zero values on the diagonal, and zeros elsewhere) can model the data well, but K-means and a mixture of spherical Gaussians cannot.
3. **8 Points.** Draw a data set where a mixture of 3 Gaussians with unrestricted covariance matrices can model the data well, but K-means and a mixture of diagonal Gaussians cannot.



(a) Aggregation (K=7) (b) Bridge (K=2) (c) Compound (K=6) (d) Flame (K=2)



(e) Jain (K=2) (f) Spiral (K=3) (g) TwoDiamond (K=2)

Figure 1: Groundtruth clustering results of 7 toy datasets.

Problem 4 (45 Points): Implementing K-Means and Spectral Clustering

Given a number of 7 toy datasets (in `toydata.zip`). Each dataset contains a number of clusters as shown in Figure 1 and your task is to find these clusters using your implemented clustering algorithms.

1. **20 Points. Implementing Lloyd's K-means:** Your submitted function should be `function [label] = my_kmeans(data, K)`, where `label` returns the N-dimensional clustering result, where N is the total number of data points. `data` is with size $N \times d$ and K is the number of (known) clusters. To initialize, randomly select K samples to initialize your cluster centroids. Iterate your algorithm until convergence. Use **Euclidean distance** as the distance measure. Name your file `my_kmeans.py`.
2. **20 Points. Implementing Spectral Clustering:** Your submitted function should be `function [label] = my_spectralclusting(data, K, sigma)`, where `label`, `data` and K are the same as above and `sigma` is the bandwith for **Gaussian kernel** used in spectral clustering. You will see `sigma` is important for your clustering performance. Adjust it case-by-case for every toy dataset to output the best results. Name your file `my_spectralclusting.py`.
3. **5 Points.** Compare your spectral clustering results with k-means. It is natural that on certain hard toy example, both method won't generate perfect results. In your report, briefly analyze what is the advantage or disadvantage of spectral clustering over k-means. Why it is the case? (You do not need to mathematically prove it but just need to give answers in your own language.)

Remarks: Write a file named `Run_clustering.py` at top level to give the clustering results. In particular, your code should:

1. Load all the mat file data and generate clustering labels with your k-means and spectral clustering codes. In the mat files, "D" is the data matrix and "L" is the ground truth label matrix. To run your clustering algorithms, you cannot use "L" as they only serve as a reference.

2. Show your clustering results in the solution (if you use word or latex) and indicate which methods (k-means or spectral clustering) generated them.

3. It is NOT allowed to use and refer to any external codes of k-means and spectral clustering. It is not hard to implement them (about 20 lines of code).

4. The default programming language is python, but it is fine for you to choose any other languages.