Assignment 5

1. (25 Points) Given a set of 5 samples

$$X = \begin{bmatrix} 0 & 0 & 1 & 5 & 5 \\ 2 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Try k-means clustering algorithm to cluster the samples to 2 classes.

2. (25 Points) Suppose there are three coins, denoted A, B, and C. The probabilities of these coins coming up heads are π, p and q. Conduct the following coin toss test. First, toss coin A and select coin B or coin C according to its result, with coin B being selected for heads and coin C for tails. Then toss the selected coin, with the result recorded as 1 for heads and 0 for tails. Repeat the test n times independently (here, n = 10). The observation results are as follows:

$$1, 1, 0, 1, 0, 0, 1, 0, 1, 1$$

Suppose that only the result of the coin toss can be observed, but not the process of tossing. The question is how to estimate the probability that all three coins will come up heads, i.e., to find the maximum likelihood estimation of the model parameters $\theta = (\pi, p, q)$.
(Assuming that the initial value of the model parameter is $\pi^{(0)} = 0.46$, $p^{(0)} = 0.55$, $q^{(0)} = 0.67$, you can use python to calculate the results).

Note: In addition to submitting the **formulas and answers**, you are also required to submit the **.py/.ipynb code**.

3. (25 Points) With the known observation data −67, −48, 6, 8, 14, 16, 23, 24, 28, 29, 41, 49, 56, 60, 75, try to estimate the parameters $(\alpha_0, \mu_0, \sigma_0, \alpha_1, \mu_1, \sigma_1)$ of the two-component Gaussian mixture model.

Note: In addition to submitting the **formulas and answers**, you are also required to submit the **.py/.ipynb code**.

4. (25 Points) Given a set of 5 samples with 4 features

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 5 & 6 & 7 \\ 1 & 4 & 2 & 3 \\ 5 & 3 & 2 & 1 \\ 8 & 1 & 2 & 2 \end{bmatrix}$$

Use PCA algorithm to reduce the samples to 2 dimensions.

Note: In addition to submitting the **formulas and answers**, you are also required to submit the **.py/.ipynb code**.