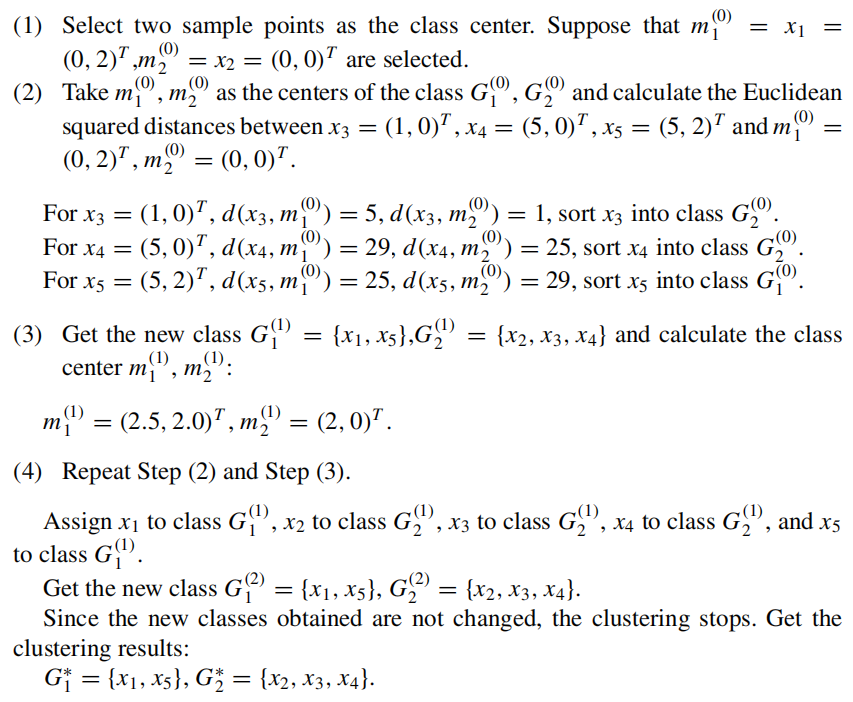
Assignment 5

1. Given a set of 5 samples

X =

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

Answer:



1. 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

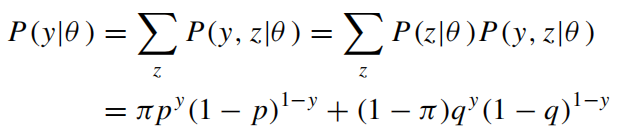
Suppose that only the result of the coin toss can be observed, but not the process of tossing. The question is how to estimate the probability that all three coins will come up heads, i.e., to find the maximum likelihood estimation of the model parameters θ = (π, p, q).

(Assuming that the initial value of the model parameter is π(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**.

Answer:

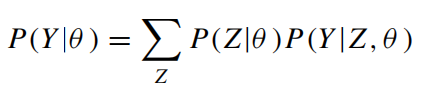
The three-coin model can be written as

** (1)

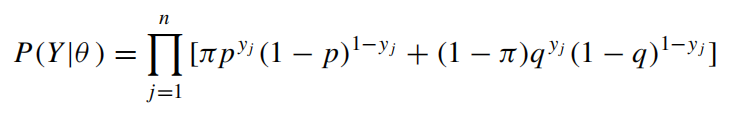
Here, the random variable *y* is the observed variable, indicating that the outcome of a trial observation is 1 or 0; the random variable *z* is the hidden variable, indi- cating the unobserved outcome of tossing coin A; and *θ (π, p, q)* is the model parameter. This model is the generative model for the above data. Note that the data for the random variable *y* is observable, and the data for the random variable *z* is unobservable.

=

Denote the observation data as *Y* = *(Y*1*, Y*2*,... Yn)*T and the un-observation data as *Z* = *(Z*1*, Z*2*,... Zn)*T, then the likelihood function of the observation data is

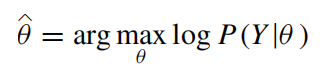
**  (2)

i.e.,

 (3)

Consider finding the maximum likelihood estimate of the model parameters *θ (π, p, q)*, i.e.,

=

**  (4)

This problem has no analytical solution and can only be solved by an iterative method. The EM algorithm is an iterative algorithm that can solve this problem.

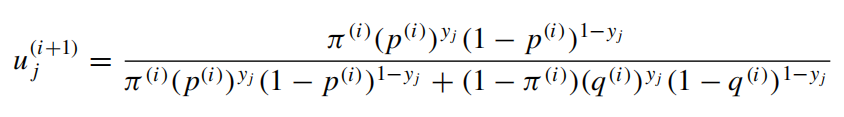
The EM algorithm for the above problem is given below, with its derivation being omitted. The EM algorithm first selects the initial values of the parameters as *θ (*0*)*

=

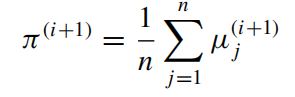
*(π (*0*), p(*0*), q(*0*))*, and then iterates through the following steps to compute the esti- mated value of the parameters until convergence. The estimated value of the param- eter in the *i* - th iteration is *θ (i ) (π (i), p(i), q(i))*. The *(i* 1*)th* iteration of the EM algorithm is as follows:

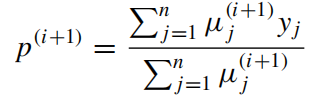
= +

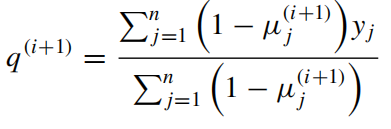
E-step: Compute the probability that observation *y j* comes from tossing coin B under model parameters π*(i), p(i), q(i)*

 (5)

M-step: Compute the new estimated value of the model parameters

 (6)

 (7)

 (8)

Conduct the numerical calculation. Assume that the initial value of the model parameter is

*π (*0*)* = 0.46*, p(*0*)* = 0*.*55*, q(*0*)* = 0*.67*

According to Eq. (5), we have

Using the iterative Eqs. (6), (7) and (8), we get

Similarly, we have

Therefore, the maximum likelihood estimation of the model parameters is:

1. 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 () of the two-component Gaussian mixture model.

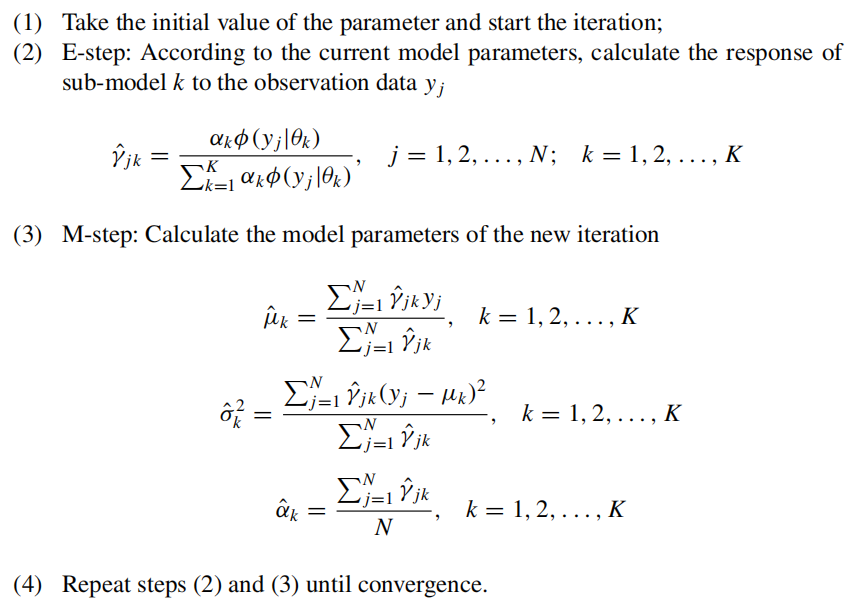
**Note**: In addition to submitting the **formulas and answers**, you are also required to submit the **.py/.ipynb code**.

Answer:

(*the EM algorithm for parameter estimation of Gaussian mixture model*)

Input: Observation data *y*1*, y*2*,..., yN* ,Gaussian mixture model;

Output: Gaussian mixture model parameters.



First, we use sklearn's GaussianMixture to compute the 6 parameters, then implement the EM algorithm for the Gaussian Mixture Model to obtain the final result as: