

ML-As-5

Q1

Given a set of 5 samples:

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

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

Consider Data point: $(0, 0), (5, 0)$

Data Point	Distance to $(0, 0)$	Distance to $(5, 0)$	Cluster
$(0, 2)$	2	$\sqrt{29}$	$(0,0)$
$(0, 0)$	0	5	$(0,0)$
$(1, 0)$	1	4	$(0,0)$
$(5, 0)$	5	0	$(5,0)$
$(5, 2)$	$\sqrt{29}$	2	$(5,0)$

Data Point	Distance to $(\frac{1}{3}, \frac{2}{3})$	Distance to $(5, 1)$	Cluster
$(0, 2)$	$\frac{\sqrt{17}}{3}$	$\sqrt{26}$	$(0,0)$
$(0, 0)$	$\frac{\sqrt{5}}{3}$	$\sqrt{26}$	$(0,0)$
$(1, 0)$	$\frac{2\sqrt{2}}{3}$	$\sqrt{17}$	$(0,0)$
$(5, 0)$	$\frac{10\sqrt{2}}{3}$	1	$(5,0)$
$(5, 2)$	$\frac{2\sqrt{53}}{3}$	1	$(5,0)$

The cluster does not change after the second iteration. The final cluster assignments are:

Class-1: $(0, 0), (0, 2), (1, 0)$

Class-2: $(5, 0), (5, 2)$

Q2

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).

The likelihood of observing the data is:

$$P(x_i) = \pi \cdot P_B(x_i) + (1 - \pi) \cdot P_C(x_i) \quad (2)$$

$$\ell(\theta) = \sum_{i=1}^n \ln (\pi \cdot p^{x_i} (1 - p)^{1-x_i} + (1 - \pi) \cdot q^{x_i} (1 - q)^{1-x_i}). \quad (3)$$

$$\gamma_{B,i} = \frac{\pi \cdot P_B(x_i)}{\pi \cdot P_B(x_i) + (1 - \pi) \cdot P_C(x_i)} \quad (4)$$

$$\gamma_{C,i} = \frac{(1 - \pi) \cdot P_C(x_i)}{\pi \cdot P_B(x_i) + (1 - \pi) \cdot P_C(x_i)} \quad (5)$$

$$P_B(x_i) = p^{x_i} (1 - p)^{1-x_i} \quad (6)$$

$$P_C(x_i) = q^{x_i} (1 - q)^{1-x_i} \quad (7)$$

$$\pi = 0.46, p = 0.53, q = 0.65 \quad (8)$$

$$p(\text{get 3 heads}) = 0.46 \times 0.53 \times 0.65 = 0.16 \quad (9)$$

Q3

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.

Initialization:

- Randomly initialize the parameters $(\alpha_0, \mu_0, \sigma_0, \alpha_1, \mu_1, \sigma_1)$.
- The weights α_0 and α_1 must satisfy $\alpha_0 + \alpha_1 = 1$.

E-step (Expectation step):

Compute the responsibility $r_{i,k}$ for each data point x_i belonging to the k -th Gaussian component:

$$r_{i,k} = \frac{\alpha_k \cdot \mathcal{N}(x_i \mid \mu_k, \sigma_k^2)}{\sum_{j=0}^1 \alpha_j \cdot \mathcal{N}(x_i \mid \mu_j, \sigma_j^2)} \quad (10)$$

where $\mathcal{N}(x \mid \mu, \sigma^2)$ is the probability density function of the Gaussian.

M-Step (Maximization step):

Update the parameters using the responsibilities:

$$\begin{aligned} \blacksquare \alpha_k &= \frac{1}{n} \sum_{i=1}^n r_{i,k} \\ \blacksquare \mu_k &= \frac{\sum_{i=1}^n r_{i,k} x_i}{\sum_{i=1}^n r_{i,k}} \\ \blacksquare \sigma_k^2 &= \frac{\sum_{i=1}^n r_{i,k} (x_i - \mu_k)^2}{\sum_{i=1}^n r_{i,k}} \end{aligned}$$

Iterate:

- Repeat the E-step and M-step until the parameters converge or the change is below a small threshold.

Component 0:

$$\alpha_0 = 0.1332, \mu_0 = -57.51, \sigma_0 = 9.50$$

Component 1:

$$\alpha_1 = 0.8668, \mu_1 = 32.98, \sigma_1 = 20.72$$