CO553 - Introduction to Machine Learning: Unsupervised Learning

Prepared by Josiah Wang

Autumn 2020/2021

1 Questions

Here are some practical exercises for you to improve your understanding of clustering and probability density estimation.

1. In the table below, you are given a dataset of 10 examples, where each sample comprises two variables x_1 and x_2.

i	1	2	3	4	5	6	7	8	9	10
$x_1^{(i)}$	1.23	0.83	0.23	1.51	-1.09	-0.50	-0.08	1.49	-0.20	2.26
$x_2^{(i)}$	0.11	-0.59	2.06	1.35	0.53	1.01	0.25	1.83	-0.77	0.88

Assume that the K-means algorithm is currently executing (assume K=2). It has just completed an update step. The mean of the two clusters has been computed as $\mu_1 = \begin{bmatrix} 0.06 \\ 0.37 \end{bmatrix}$ and $\mu_2 = \begin{bmatrix} 1.75 \\ 1.35 \end{bmatrix}$ respectively. The algorithm finds that the cluster means have not yet converged.

   (a) **Assignment step:** Compute the cluster assignments for all 10 examples with the latest cluster means μ_1 and μ_2. Assume that the K-means algorithm uses **Euclidean distance** as its distance measure.

   (b) **Update step:** Compute the new mean of the two clusters using the cluster assignments that you have computed.

2. In the table below, you are given a dataset with five training examples.

i	1	2	3	4	5
$x^{(i)}$	7.42	2.28	3.45	7.17	1.75

Using this training dataset, compute the probability density $p(x = 3.13)$ using Kernel Density Estimation, assuming a bandwidth of h=1. The probability density function for a Kernel Density Estimator is defined as: $p(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi h^2}} \exp^{\left(-\frac{(x-x^{(i)})^2}{2h^2}\right)}$.

3. Using the same dataset as in Q2 above, fit the parameters of a Gaussian distribution by computing its mean μ and variance σ^2. Then compute the probability density $p(x = 3.13|\mu, \sigma^2)$ given the Gaussian distribution that you have fitted.

The Gaussian distribution is defined as $\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where the mean is $\mu = \frac{1}{N} \sum_{i=1}^{N} x^{(i)}$ and the variance is $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \mu)^2$.

4. Consider the dataset in the table below with five training examples.

i	1	2	3	4	5
$x^{(i)}$	5.92	2.28	3.85	5.17	1.75

Assume that a two-component Gaussian Mixture Model (GMM) has been initialised as follows.

k	π_k	μ_k	σ_k^2
1	0.5	3.34	1.0
2	0.5	6.12	1.0

Perform one iteration of the Expectation-Maximisation (EM) algorithm.

(a) **Expectation step**: Compute the responsibilities r_{ik} for each example i and component k.

(b) **Maximisation step**: Compute the mean μ_k, variance σ_k^2 and mixing proportion π_k for each component k given the responsibilities you computed earlier in the E-step.

Here are some equations which you may find useful:

- $\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $\mu_k = \frac{1}{\sum_{j=1}^{N} r_{jk}} \sum_{i=1}^{N} r_{ik} x^{(i)}$
- $\sigma_k^2 = \frac{1}{\sum_{j=1}^{N} r_{jk}} \sum_{i=1}^{N} r_{ik} (x^{(i)} - \mu_k)^2$
- $\pi_k = \frac{1}{N} \sum_{i=1}^{N} r_{ik}$

5. Suppose the parameters for the GMM in Q4 are fitted as follows after convergence:

k	π_k	μ_k	σ_k^2
1	0.61	2.65	0.85
2	0.39	5.55	0.14

Compute the probability density $p(x = 3.13|\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ given the GMM above.