Problem 1. *K-means and Gaussian Mixture Model*
The Gaussian mixture model can be written as the superposition of Gaussians in the form

$$p(\mathbf{x}_i) = \sum_{k=1}^{K} p_k \, \phi(\mathbf{x}_i \mid \mu_k, \mathbf{\Sigma}_k) \,,$$

where ϕ is the normal density with mean μ_k and covariance matrix $\mathbf{\Sigma}_k$, and p_k the probability that the i-th observation $\mathbf{x}_i$ belongs to class k.

(i) Derive the EM algorithm for the Gaussian mixture model.

The K-means algorithm partitions the dataset into K clusters, where K is supposed fixed in advance. For each datapoint $\mathbf{x}_i$, we assign a binary variable π_{ik} describing which cluster the variable belongs to,

$$\pi_{ik} = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ belongs to cluster } k \\ 0 & \text{otherwise.} \end{cases}$$

The goal is to select $\pi := \{\pi_{ik}\}$ and K centers z_k such that

$$\mathcal{C}(z_1, \ldots, z_K, \pi) = \sum_{k=1}^{K} \sum_{i=1}^{n} \pi_{ik} \, ||\mathbf{x}_i - z_k||^2$$

is minimized.

(ii) Show that this can be achieved using a two-step iterative procedure, where

Step *(a)* Assign $\mathbf{x}_i$ to the nearest μ_k

Step *(b)* Update μ_k using

$$\mu_k = \frac{\sum_i \pi_{ik} \, \mathbf{x}_i}{\sum_i \pi_{ik}} \,.$$

(iii) Consider a Gaussian mixture model with covariance matrices given by $\epsilon \, \mathbf{I}$, where ϵ is a variance parameter shared by all components. Show that in the limit $\epsilon \to 0$, the EM re-estimation for the Gaussian means μ_k reduces to the K-mean result. Then show that in this limit, maximising the expected complete log-likelihood is equivalent to minimising $\mathcal{C}(z_1, \ldots, z_K, \pi)$.

Problem 2. *EM algorithm for binomial count data*
Consider a set of n observations $\mathcal{L}_n = \{x_1, \ldots, x_n\}$, where each $x_i \in \{0, 1, 2, \ldots\}$. Assume that observation x_i belongs to category k with probability π_k, with $k = 1, \ldots, K$, and that given x_i belongs to category k,

$$x_i \sim Bi(n_i, \, p_k) \,.$$

We do not know the category label of each observation, but we assume that the n_i are known. Write down the EM algorithm estimating the parameters $\pi_1, \ldots, \pi_K, p_1, \ldots, p_K$.

Problem 3. *EM algorithm for censored data*
Suppose that $x_1, \ldots, x_n$ are independent truncated observations of a normally distributed random variable. Specifically, each x_i is a realisation of a generic X, where

$$X = \min(X^*,\, a)\,, \quad \text{and } X^* \sim \mathcal{N}(\theta, 1)\,.$$

We suppose $a > 0$ known, and we wish to estimate θ based solely on observations $x_1, \ldots, x_n$. We denote by $x_1^*, \ldots, x_n^*$ the associated partially observed variables.

(i) To simplify notation, denote the first $m \leq n$ observations $x_1 = x_1^*, \ldots, x_m = x_m^*$ as uncensored, and the remaining $(n-m)$ as censored. The censored observations are treated as latent variables. We denote them by $z_{m+1} = x_{m+1}^*, \ldots, z_n = x_n^*$ (instead of $z_{m+1}, \ldots, z_n$, you observe $n-m$ times the value a, as $x_{m+1} = \ldots = x_n = a$). Ignoring constants independent of θ, show that the complete log-likelihood can be written

$$\ell(\theta | x_1, \ldots, x_m, z_{m+1}, \ldots, z_n) = -\frac{1}{2}\sum_{i=1}^{m}(x_i - \theta)^2 - \frac{1}{2}\sum_{i=m+1}^{n}(z_i - \theta)^2\,.$$

(ii) Deduce from *(i)* the expression of the likelihood,

$$f(\theta | x_1 \ldots, x_n) = \frac{1}{(2\pi)^{m/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}(x_i - \theta)^2\right\} (1 - \Phi(a - \theta))^{n-m}\,.$$

(iii) Show that the latent variables z_i, for $i = m+1, \ldots, n$, have conditional density

$$k(z_i | X_i = a, \theta) = \frac{\varphi(z_i - \theta)}{1 - \Phi(a - \theta)}\,, \quad a \leq z_i\,,$$

and zero elsewhere, where φ and Φ denote the standard normal density and distribution, respectively.

(iv) We compute the E-step of the EM algorithm. Show that

$$\mathbf{E}(Z | \theta^{(m)}) = \theta^{(m)} + \frac{\varphi(a - \theta^{(m)})}{1 - \Phi(a - \theta^{(m)})}\,,$$

and write down the expression of $Q(\theta, \theta^{(m)})$.

(v) Show that the M-step returns the following update

$$\theta^{(m+1)} = \frac{m}{n}\bar{x} + \frac{n-m}{n}\left(\theta^{(m)} + \frac{\varphi(a - \theta^{(m)})}{1 - \Phi(a - \theta^{(m)})}\right)\,,$$

where $\bar{x} = m^{-1}\sum_{i=1}^{m} x_i$.