Bernoulli Likelihood with EM

1 Bernoulli Distribution

Observed data: $X_1, X_2, \ldots, X_N \sim$ i.i.d. Bernoulli$(X \mid \mu)$. Each X_i is a vector of 0's and 1's.

$$X_i = [x_{i1}, x_{i2}, \ldots, x_{iD}] \quad x_{id} = \{0, 1\}.$$

So, N observations and D features.

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{pmatrix}.$$

Density:

$$\text{Bernoulli}(X_i \mid \mu) = \prod_{d=1}^{D} \mu_d^{x_{id}} (1 - \mu_d)^{1 - x_{id}}.$$

Now, let's consider a finite mixture of K Bernoulli distributions. Then, its density function can be expressed as follows:

$$\mathbb{P}(X_i \mid \mu, \pi) = \sum_{k=1}^{K} \pi_k \mathbb{P}(X_i \mid \mu_k),$$

where $\mu = [\mu_1, \ldots, \mu_K], \pi = [\pi_1, \ldots, \pi_K]$, and $\mathbb{P}(X_i \mid \mu_k) = \prod_{d=1}^{D} \mu_{kd}^{x_{id}} (1 - \mu_{kd})^{1 - x_{id}}$.

2 Inference with EM

Given the observed data $X = [X_1, X_2, \ldots, X_N]$, we can write the log-likelihood function for this model as follows:

$$\log \mathbb{P}(X \mid \mu, \pi) = \sum_{i=1}^{N} \log \left[\sum_{k=1}^{K} \pi_k \mathbb{P}(X_i \mid \mu_k) \right].$$

We cannot solve it for the maximum likelihood solution in closed form due to the summation within the logarithm. We now derive the EM algorithm for mixture of Bernoulli distributions. Let the latent variable $z_i = [z_{i1}, \ldots, z_{iK}]$ be a binary K-dimensional variable with a single component equal to 1 and the rest equal to 0, indicating the cluster assignment of X_i.

Now, the expectation of the log-likelihood is given by:

$$\mathbb{E}[\log \mathbb{P}(X, Z \mid \mu, \pi)] = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \left[\log \pi_k + \sum_{d=1}^{D} x_{id} \log \mu_{kd} + (1 - x_{id}) \log(1 - \mu_{kd}) \right],$$

where $r_{ik} = \mathbb{E}[z_{ik}]$ is the probability that X_i belongs to cluster k.

E-step: Compute the probabilities of cluster assignments.

$$\begin{aligned} r_{ik} &= \frac{\pi_k \mathbb{P}(X_i \mid \mu_k)}{\sum_{j=1}^{K} \pi_j \mathbb{P}(X_i \mid \mu_j)} \\ &\propto \pi_k \prod_{d=1}^{D} \left[\mu_{kd}^{x_{id}} (1-\mu_{kd})^{(1-x_{id})} \right]. \end{aligned}$$

M-step: Update parameters μ and π given r.

$$\begin{aligned} \mu_k &= \frac{1}{N_k} \sum_{i=1}^{N} r_{ik} X_i \\ \pi_k &= \frac{N_k}{N}, \end{aligned}$$

where $N_k = \sum_{i=1}^{N} r_{ik}$.