CS5691

①

Derivation of EM algorithm

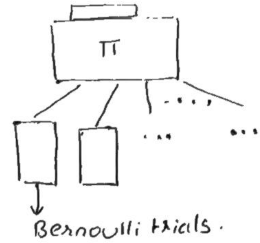

Fig. Schematic of the Model

- Notation

* Data: $X = \{x_1 \dots x_n\}$, $x_i \in \mathbb{R}^d \quad \forall \; 1 \le i \le n$

  $x_i = \{x_i^1 \dots x_i^d\} \quad x_i^j \in \mathbb{R} \quad \forall \; 1 \le j \le d$

* $\Pi = \{\pi_1 \cdots \pi_K\} \quad \pi_k \in \mathbb{R}, \; 0 \le \pi_k \le 1 \; \forall \; 1 \le k \le K$

  such that,

  $$\sum_{k=1}^{K} \pi_k = 1$$

* $$\Theta = \begin{bmatrix} p_1^1 & \cdots & p_1^d \\ \vdots & & \vdots \\ p_K^1 & \cdots & p_K^d \end{bmatrix} \quad \Theta \in \mathbb{R}^{K \times d}, \quad 0 \le p_k^j \le 1 \quad \forall \; 1 \le j \le d, \; 1 \le k \le K$$

Assume $\{x_1 \cdots x_n\}$ drawn from i.i.d Bernoulli trials (each x_i^j is a bernoulli trial)

- Likelihood

$$\mathcal{L}(\Theta; X) = \prod_{i=1}^{n} p(x_i; \Theta) \qquad (\text{By i.i.d})$$

$$= \prod_{i=1}^{n} \left[\sum_{k=1}^{K} \pi_k \, p(x_i; \Theta_k) \right] \qquad \left(\begin{array}{l} \text{By thm of total Probability} \\ \text{and } \Theta_K \text{ is } k^{th} \text{ row of } \Theta \end{array} \right)$$

$$= \prod_{i=1}^{n} \left[\sum_{k=1}^{K} \pi_k \left(\prod_{j=1}^{d} (p_k^j)^{x_i^j} (1 - p_k^j)^{1 - x_i^j} \right) \right]$$

[Note: $p(x_i; \Theta_K) = \prod_{j=1}^{d} (p_K^j)^{x_i^j} (1 - p_K^j)^{1 - x_i^j}$ [Bernoulli trials]]

* Introduce $\lambda_K^i \quad 1 \le i \le n, \; 1 \le k \le K$

$$\log \mathcal{L}(\Theta; X) = \sum_{i=1}^{n} \log \left[\sum_{k=1}^{K} \lambda_K^i \left(\frac{\pi_k \prod_{j=1}^{d} (p_K^j)^{x_i^j} (1 - p_K^j)^{1 - x_i^j}}{\lambda_K^i} \right) \right]$$

* log is concave, applying Jensen's inequality.

$$\text{mod} \cdot \log \mathcal{L}(\Theta; X) = \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_K^i \log \left(\frac{\pi_K \, p(x_i; \Theta_K)}{\lambda_K^i} \right)$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_K^i \left[\log \left(\frac{\pi_K}{\lambda_K^i} \right) + \sum_{i=1}^{d} \left(x_i^j \log p_K^j + (1 - x_i^j) \log (1 - p_K^j) \right) \right]$$

Here

$$\log \mathcal{L}(\Theta) \geq \text{mod} \log \mathcal{L}(\Theta, \lambda) \qquad (\text{Since log is Concave})$$

Note: Constraint on λ is $\sum_{k=1}^{K} \lambda_K^i = 1 \quad \forall \; i \quad 0 \le \lambda_K^i \le 1 \quad \forall \; K$

- Obtaining Maximum Likelyhood estimators (Θ, Π)

$$\frac{\partial \, \text{mod} \log L(\Theta)}{\partial P_k^j} = \sum_{i=1}^{n} \lambda_k^i \left(\frac{x_i^j}{P_k^j} - \frac{(1-x_i^j)}{1-P_k^j} \right) = 0$$

$$\therefore \quad \sum_{i=1}^{n} \left(\lambda_k^i \left(x_i^j - x_i^j P_k^j - P_k^j + P_k^j x_i^j \right) \right) = 0$$

$$\Rightarrow \quad \boxed{P_k^j = \frac{\sum_{i=1}^{n} \lambda_k^i x_i^j}{\sum_{i=1}^{n} \lambda_k^i}} \quad \forall\, 1 \le j \le d, \quad 1 \le k \le K \quad \text{———} \; (1)$$

Now,

$$\max_{\substack{\pi_1 \dots \pi_K \\ \text{s.t.} \sum_{k=1}^{K} \pi_k = 1}} \text{mod} \log L(\Theta) \equiv \min_{\substack{\pi \in \mathbb{R}^K \\ \text{s.t.} \sum_{k=1}^{K} \pi_k = 1}} - \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \left(\log\left(\frac{\pi_k}{\lambda_k^i}\right) + \log P(x_i, \Theta_k) \right)$$

Reparametrize $\pi_k = e^{\gamma_k}$, $\sum e^{\gamma_k} = 1$

$$\nabla(-\text{mod} \log(\Theta, \lambda))_k = -\mathcal{L} e^{\gamma_k} \quad (\mathcal{L} \rightarrow \text{Lagrange multiplier})$$

$$\therefore \quad \sum_{i=1}^{n} \lambda_k^i = +\mathcal{L} e^{\gamma_k} \Rightarrow \pi_k = \frac{\sum_{i=1}^{n} \lambda_k^i}{\mathcal{L}}$$

now $$\sum_{k=1}^{K} \pi_k = 1 = \frac{\sum_{i=1}^{n} \sum_{k=1}^{K} (\lambda_k^i)}{\mathcal{L}} = \frac{n}{\mathcal{L}} \Rightarrow n = \mathcal{L}$$

$$\therefore \quad \boxed{\pi_k = \frac{\sum_{i=1}^{n} \lambda_k^i}{n}} \quad \forall\, 1 \le k \le K \quad \text{———} \; (2)$$

Now Maximize mod $\log L(\Theta)$ over λ with (Θ, π) as constants.

$$\max_{\substack{\lambda_k^i \dots \\ \text{s.t.} \sum_{i=1}^{K} \lambda_k^i = 1}} \sum_{k=1}^{K} \lambda_k^i \left(\log\left(\frac{\pi_k}{\lambda_k^i}\right) + \log(P(x_i, \Theta_k)) \right) \equiv \min_{\substack{\lambda_1^i \dots \lambda_K^i \\ \text{s.t.} \sum_{i=1}^{K} \lambda_k^i = 1}} - \sum_{k=1}^{K} \lambda_k^i \left(\log\left(\frac{\pi_k}{\lambda_k^i}\right) + \log(P(x_i, \Theta_k)) \right)$$

Reparametrize $\lambda_k^i = e^{\gamma_k}$

$$\nabla\left(-\sum_{k=1}^{K} \lambda_k^i \left(\log\left(\frac{\pi_k}{\lambda_k^i}\right) + \log(P(x_i; \Theta_k)) \right) \right)_k = -\mathcal{L} e^{\gamma_k} \quad (\mathcal{L} \text{ is lagrange multiplier})$$

$$\therefore \quad e^{\gamma_k} \left(\log \pi_k + \log(P(x_i; \Theta_k)) - \gamma_k - 1 \right) = \mathcal{L} e^{\gamma_k}$$

$$\Rightarrow \quad \gamma_k = \log \pi_k P(x_i; \Theta_k) - \mathcal{L} - 1 = \log \lambda_k^i$$

$$\therefore \quad (e^{-\mathcal{L}-1}) \pi_k P(x_i; \Theta_k) = \lambda_k^i \quad \text{(take exp on both sides)}$$

Now $$\sum_{k=1}^{K} \lambda_k^i = 1 \quad \therefore \quad e^{-\mathcal{L}-1} = \frac{1}{\sum_{k=1}^{K} \pi_k P(x_i; \Theta_k)} \Rightarrow \boxed{\lambda_k^i = \frac{\pi_k P(x_i, \Theta_k)}{\sum_{l=1}^{K} \pi_l P(x_i; \Theta_l)}} \text{—} (3)$$