ML2 Home Assignment 5

id: 12179078

Volodymyr Medentsiy

PROBLEM 1

GMM: $p(x) = \sum_{k=1}^{K} \pi_k N(x | \mu_k, \Sigma_k)$

1. Given: $E_{posterior}(\ln p(X, Z | \mu, \Sigma, \pi)) = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \cdot$

$\cdot (\ln \pi_k + \ln N(x_n | \mu_k, \Sigma_k))$.

Derive update rules for μ, Σ, π.

1) To derive update rules I'll find gradients with respect to parameters μ_k, Σ_k and π_k.

$$E_{posterior}(\ln p(X, Z | \mu, \Sigma, \pi)) = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \Big(\ln \pi_k - \frac{D}{2} \ln 2\pi -$$

$$- \frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) \Big),$$ assuming $x_n \in \mathbb{R}^D$

$$\frac{\partial E_{posterior}}{\partial \mu_k}(\ln p(X, Z | \mu, \Sigma, \pi)) = \sum_{n=1}^{N} \gamma(z_{nk}) \Sigma_k^{-1} (x_n - \mu_k) = 0 \,\Big|\, \times \Sigma_k$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(z_{nk}) \cdot \mu_k = \sum_{n=1}^{N} \gamma(z_{nk}) x_n \Rightarrow \mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})} =$$

$$= \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{N_k} \quad \text{with} \quad N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

2)
$$\frac{\partial E_{posterior}}{\partial \Sigma_k}(\ln p(X, Z | \mu, \Sigma, \pi)) = \sum_{n=1}^{N} \gamma(z_{nk}) \Big(- \frac{\partial}{\partial \Sigma_k} \ln |\Sigma_k| -$$

$$- \frac{1}{2} \frac{\partial}{\partial \Sigma_k} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) \Big) =$$

To derive this gradient I will use the following identities, taken from the « The Matrix Cookbook »

$$\frac{\partial \ln|\det X|}{\partial X} = (X^T)^{-1}$$, in case of X symmetric then $(X^T)^{-1} = X^{-1}$.

$$\frac{\partial a^T X^{-1} b}{\partial X} = -(X^T)^{-1} a b^T (X^T)^{-1}$$

$$\Leftrightarrow \sum_{n=1}^{N} \gamma(z_{nk}) \left(-\frac{1}{2} \Sigma_k^{-1} + \frac{1}{2} \Sigma_k^{-1} (x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1} \right) = 0 \Big| \times \Sigma_k$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(z_{nk}) \cdot I = \sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{-1}$$, where I is an identity matrix

$$\Rightarrow \Sigma_k^{-1} = N_k \cdot I \left(\sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T \right)^{-1} \Rightarrow$$

$$\Rightarrow \Sigma_k = \frac{1}{N_k} \cdot \sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T$$

3) When deriving an update rule for π_k we need to keep in mind that π_k is subject to constraint : $\sum_{k=1}^{K} \pi_k = 1$. Thus we introduce Lagrange multiplier λ and our objective function changes to:

$$E_{posterior}(\ln p(X, Z | \mu, \Sigma, \pi)) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) \longrightarrow \max_{\pi}$$

$$\frac{\partial \left(E_{posterior}(\ln p(X, Z | \mu, \Sigma, \pi)) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) \right)}{\partial \pi_k} =$$

$$= \sum_{n=1}^{N} \gamma(z_{nk}) \frac{1}{\pi_k} + \lambda = 0 \Rightarrow \lambda \pi_k = -N_k \Big| \text{ sum over } k \Rightarrow$$

$$\Rightarrow \quad \lambda \underbrace{\sum_{k=1}^{K} \pi_k}_{1} = -N \quad \Rightarrow \lambda = -N.$$

Thus $\quad \pi_k = \dfrac{N_k}{N}.$

2. if $\Sigma_k = \Sigma \ \forall k$ then update rule for $\Sigma_k = \Sigma$ will change!

$$\frac{\partial}{\partial \Sigma} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left(-\frac{1}{2} \ln |\Sigma| - \frac{1}{2} (x_n - \mu_k)^T \Sigma^{-1} (x_n - \mu_k) \right) =$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left(-\frac{1}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} (x_n - \mu_k)(x_n - \mu_k)^T \Sigma^{-1} \right) = 0 \quad | \times \Sigma$$

$$\Rightarrow \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left(-\frac{1}{2} \cdot I + \frac{1}{2} (x_n - \mu_k)(x_n - \mu_k)^T \Sigma^{-1} \right) = 0$$

$$\Rightarrow \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \cdot I = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T \Sigma^{-1}$$

$$\Rightarrow \Sigma^{-1} = \frac{N}{\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T}$$

$$\Rightarrow \Sigma = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T$$

PROBLEM 2

EM to maximize the posterior $p(\Theta|X)$.

1. $p(\Theta|X) = \frac{p(X|\Theta)p(\Theta)}{p(X)} \propto p(X|\Theta)p(\Theta)$

Thus the objective function changes to:

$\ln(p(X|\Theta)p(\Theta)) = \ln p(X|\Theta) + \ln p(\Theta) = $ [Bishop eq. 9.70 - 9.72] $=$

$= L(q,\Theta) + KL(q\|p) + \ln p(\Theta) \geq L(q,\Theta) + \ln p(\Theta) = \tilde{L}(q,\Theta)$

During E-step we maximize $\tilde{L}(q,\Theta)$ with respect to q:

$$q(z) = \arg\max_q \tilde{L}(q,\Theta) = \arg\max_q (L(q,\Theta) + \ln(p(\Theta))) =$$

$$= \arg\max_q L(q,\Theta)$$

So because $\ln p(\Theta)$ does not depend on q the E-step does not change.

2. During M-step we maximize $\tilde{L}(q,\Theta)$ with respect to Θ:

$$\Theta^{new} = \arg\max_\Theta \tilde{L}(q,\Theta) = \arg\max_\Theta (L(q,\Theta) + \ln p(\Theta)) =$$

$= $ [using Bishop's 9.74] $= \arg\max_\Theta \Big(\sum_Z p(Z|X,\Theta^{old}) \ln p(X,Z|\Theta) -$

$$- \underbrace{\sum_Z p(Z|X,\Theta^{old}) \ln p(Z|X,\Theta^{old})}_{\text{does not depend on } \Theta} + \ln p(\Theta)\Big) =$$

$$= \underset{\Theta}{\operatorname{argmax}}\left(\sum_{Z} p(Z|X,\Theta^{old}) \ln p(X,Z|\Theta) + \ln p(\Theta)\right).$$

PROBLEM 3

Using results of the previous problem we can conclude that E-step does not change and we can use Bishop's 9.56.

Let's derive M-step:

We need to maximize with respect to μ and π:

$$E_Z \ln p(X,Z|\mu,\pi) + \ln p(\mu) + \ln p(\pi)$$

Update rule for μ_{ki}: $\dfrac{\partial\left(E_Z \ln p(X,Z|\mu,\pi) + \ln p(\mu) + \ln p(\pi)\right)}{\partial \mu_{ki}} =$

$$= \sum_{n=1}^{N} \gamma(z_{nk}) \cdot \left(x_{ni}\frac{1}{\mu_{ki}} - \frac{1-x_{ni}}{1-\mu_{ki}}\right) + \frac{\partial}{\partial \mu_{ki}} \ln \prod_{k=1}^{K}\prod_{i=1}^{D} \frac{\mu_{ki}^{a_k-1}(1-\mu_{ki})^{1-b_k}}{B(a_k,b_k)} =$$

$$= \sum_{n=1}^{N} \gamma(z_{nk}) \left(\frac{x_{ni}}{\mu_{ki}} - \frac{1-x_{ni}}{1-\mu_{ki}}\right) + \frac{a_k-1}{\mu_{ki}} - \frac{b_k-1}{1-\mu_{ki}} = 0 \quad | \times \mu_{ki}(1-\mu_{ki})$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(z_{nk})\left(x_{ni}(1-\mu_{ki}) - (1-x_{ni})\mu_{ki}\right) + (a_k-1)(1-\mu_{ki}) - (b_k-1)\mu_{ki} = 0$$

$$\Rightarrow \sum_{n=1}^{N} \gamma(z_{nk})(x_{ni}-\mu_{ki}) + a_k - 1 - \mu_{ki}a_k + \mu_{ki} - (b_k-1)\mu_{ki} = 0 \Rightarrow$$

$$\Rightarrow \mu_{ki} \cdot (-N_k - a_k - b_k + 2) = -\sum_{n=1}^{N} \gamma(z_{nk}) x_{ni} - a_k + 1$$

$$\Rightarrow \mu_{ki} = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_{ni} + a_k - 1}{N_k + a_k + b_k - 2}.$$

Update rule for π_k:

When maximizing with respect to π_k we need to introduce Lagrange multipliers, because π_k are subject to constraint $\sum_k \pi_k = 1$

$$E_z \ln p(X, Z \mid \mu, \pi) + \ln p(\pi) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right) \rightarrow \max_{\pi_k}$$

$$\frac{\partial}{\partial \pi_k}\left(E_z \ln p(X, Z \mid \mu, \pi) + \ln p(\pi) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)\right) =$$

$$= \sum_{n=1}^{N} \gamma(z_{nk}) \frac{1}{\pi_k} + \frac{\partial}{\partial \pi_k} \ln \prod_{k=1}^{K} \frac{1}{B(\alpha)} \pi_k^{\alpha_k - 1} + \lambda =$$

$$= \sum_{n=1}^{N} \gamma(z_{nk}) \frac{1}{\pi_k} + \frac{\alpha_k - 1}{\pi_k} + \lambda = 0 \quad \Big| \times \pi_k \Rightarrow$$

$$\Rightarrow \quad N_k + \alpha_k - 1 = -\pi_k \lambda \quad \Big| \sum_{k=1}^{K}$$

$$\Rightarrow \quad N + \sum_{k=1}^{K} \alpha_k - K = -\lambda \quad \Rightarrow \quad \lambda = K - N - \alpha_0, \text{ with } \alpha_0 = \sum_{k=1}^{K} \alpha_k$$

Thus $\pi_k = \dfrac{N_k + \alpha_k - 1}{N + \alpha_0 - K}$