علیرضا دهقانپور فراشاه ۹۸۱۰۱۵۵۵

سؤال یک :

۱. بله K-means یک حالت حدی از الگوریتم EM است. اگر $\Sigma = \epsilon I$ بگیریم برای هر کلاس خواهیم داشت :

$$\gamma_j^i = \frac{\pi_j \mathcal{N}(x^{(i)} \mid \mu_j, \epsilon I)}{\sum_K \pi_K \mathcal{N}(x^{(i)} \mid \mu_K, \epsilon I)} = \frac{\pi_j \exp\left(\frac{-\|x^{(i)} - \mu_j\|^2}{2\epsilon}\right)}{\sum_K \pi_K \exp\left(\frac{-\|x^{(i)} - \mu_K\|^2}{2\epsilon}\right)}$$

* اگر $\|x^{(i)} - \mu_j\|$ مینیمم باشد آن گاه با ضرب $e^{\frac{\|x^{(i)} - \mu_j\|^2}{2\epsilon}}$ در صورت و مخرج خواهیم داشت :

$$\gamma_j^i = \frac{\pi_j}{\pi_j + \sum_K \exp\left(\frac{\|x^{(i)} - \mu_j\|^2 - \|x^{(i)} - \mu_K\|^2}{2\epsilon}\right)}$$

که چون $\|x^{(i)} - \mu_K\| > \|x^{(i)} - \mu_j\|$ خواهد بود $\|x^{(i)} - \mu_j\|^2 - \|x^{(i)} - \mu_K\|^2 < 0$ خواهد بود و خواهیم داشت :

$$\lim_{\epsilon \to 0} \gamma_j^i = \frac{\pi_j}{\pi_j + \sum \exp\left(\frac{\Delta}{2\epsilon}\right)} = 1$$

زیرا $\lim_{\substack{\epsilon \to 0 \\ \Delta < 0}} e^{\frac{\Delta}{2\epsilon}} = 0$

حال γ_j^i به ازای $x^{(i)}$ هایی که به μ_j نزدیک ترند 1 و بقیه 0 خواهد شد. در EM داشتیم :

$$\mu_j = \frac{\sum_{i=1}^{n} \gamma_j^i x^{(i)}}{\sum_{i=1}^{n} \gamma_j^i} = \frac{\sum_{x^{(i)} \in C_j} x^{(i)}}{|C_j|}$$

بنابراین به همان چیزی که در K-Means داشتیم رسیدیم.

$$P(x) = P(x_a, x_b) = P(x_a|x_b)\, P(x_b)$$

$$P(x) = P(x_a|x_b)\, P(x_b) = \sum_{K} \pi_K\, P(x|K)$$

$$\Rightarrow P(x_a|x_b) = \sum_{K} \frac{\pi_K\, P(x_a, x_b|K)}{P(x_b)} = \sum_{K} \frac{\pi_K \times P(x|K)}{\sum_{K'} \pi_{K'}\, P(x_b|K')}$$

می‌توان $\dfrac{\pi_K}{\sum_{K'} \pi_{K'}\, P(x_b|K')}$ را ضریب $P(x|K)$ گرفت ولی می‌توان جلوتر رفت و ضرایب $P(x_a|x_b,K)$ را نیز بدست آورد. داریم:

$$P(x_a|x_b) = \sum_{K} \frac{\pi_K\, P(x_a, x_b|K)}{\sum_{K} \pi_{K'}\, P(x_b|K')} =$$

$$= \sum_{K} \frac{\pi_K\, P(x_b|K)\; P(x_a|x_b, K)}{\sum_{K'} \pi_{K'}\, P(x_b|K')} \Longrightarrow$$

ضرایب $P(x_a|x_b,K)$ می‌شود:

$$\pi'_K = \frac{\pi_K\, P(x_b|K)}{\sum_{K'} \pi_{K'}\, P(x_b|K')}$$

که مجموع π'_K برابر یک خواهد شد.

Scanned by CamScanner

۱.۳ برای تمامی پارامترها از ابتدا مشتق $\ln P(x|\pi,\mu,\Sigma)$ محاسبه می‌کنیم.

$$N_K := \sum_{i=1}^{N} \frac{\pi_K \mathcal{N}(x^{(i)}|\mu_K, \Sigma)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x^{(i)}|\mu_j, \Sigma)}$$

$$L(\lambda, \ell) = \sum_{i=1}^{N} \ln \sum_{j}^{K} \pi_j \, P(x^{(i)}|\mu_j, \Sigma) + \lambda\left(\sum_{j}^{K} \pi_j - 1\right)$$

$$\frac{\partial L}{\partial \pi_K} = \underbrace{\sum_{i=1}^{N} \frac{P(x^{(i)}|\mu_K, \Sigma)}{\sum_{j=1}^{K} \pi_j P(x^{(i)}|\mu_j, \Sigma)}}_{\frac{N_K}{\pi_K}} + \lambda = 0 \implies$$

$$\implies \frac{N_K}{\pi_K} = -\lambda \implies \pi_K = \frac{-N_K}{\lambda} \quad \sum_{K} \pi_K = 1 = \frac{\sum -N_K}{\lambda} =$$

$$= \frac{-N}{\lambda} = 1 \implies \lambda = -N \implies \boxed{\pi_K = \frac{N_K}{N}} \quad (1)$$

$$\frac{\partial \text{ log-likelihood}}{\partial \mu_K} = \frac{\partial \sum_{i=1}^{N} \ln \sum_{j}^{K} \pi_j P(x^{(i)}|\mu_j, \Sigma)}{\partial \mu_K} =$$

$$\sum_{i=1}^{N} \frac{\pi_K \frac{\partial}{\partial \mu_K} \frac{1}{\sqrt{2\pi}^d |\Sigma|^{1/2}} e^{-\frac{(x^{(i)}-\mu_K)^T(\Sigma^{-1})(x^{(i)}-\mu_K)}{2}}}{\sum_{j=1}^{K} \pi_j P(x^{(i)}|\mu_j, \Sigma)} =$$

$$= \sum_{i=1}^{N} \frac{\pi_K \frac{1}{\sqrt{2\pi}^d |\Sigma|^{1/2}} e^{-(x^{(i)}-\mu_K)^T(\Sigma^{-1})(x^{(i)}-\mu_K)} \times \Sigma^{-1}(x^{(i)}-\mu_K)}{\sum_{j=1}^{K} \pi_j P(x^{(i)}|\mu_j, \Sigma)} =$$

ادامه صفحه بعد

$$\sum_{i=1}^{N} \frac{\pi_K P(x^{(i)}|\mu_K,\Sigma)\,\Sigma^{-1}(x^{(i)}-\mu_K)}{\sum_j \pi_j P(x^{(i)}|\mu_j,\Sigma)} = 0 \quad \xRightarrow{\Sigma\times}$$

$$\sum_{i=1}^{N} \frac{\pi_K P(x^{(i)}|\mu_K,\Sigma)\, x^{(i)}}{\sum_j \pi_j P(x^{(i)}|\mu_j,\Sigma)} = \mu_K \sum_{i=1}^{N} \underbrace{\frac{\pi_K P(x^{(i)}|\mu_K,\Sigma)}{\sum_j \pi_j P(x^{(i)}|\mu_j,\Sigma)}}_{N_K}$$

$$\Rightarrow \boxed{\mu_K = \frac{1}{N_K} \sum_{i=1}^{N} \frac{\pi_K P(x^{(i)}|\mu_K,\Sigma)\, x^{(i)}}{\sum_j \pi_j P(x^{(i)}|\mu_j,\Sigma)}} \qquad (2)$$

$$\frac{\partial\, \text{log-likelihood}}{\partial \Sigma} = \sum_{i=1}^{N} \frac{\partial}{\partial \Sigma} \ln \sum_{j=1}^{K} \pi_j P(x^{(i)}|\pi_j,\Sigma) =$$

$$= \sum_{i=1}^{N} \frac{\sum_{j=1}^{K} \frac{\partial}{\partial \Sigma} \pi_j P(x^{(i)}|\pi_j,\Sigma)}{\sum_{j'=1}^{K} \pi_{j'} P(x^{(i)}|\pi_{j'},\Sigma)}$$

$$\frac{\partial}{\partial \Sigma} \pi_j P(x^{(i)}|\pi_j,\Sigma) =$$

ابتدا صورت را محاسبه می کنیم:

$$= \frac{\partial}{\partial \Sigma} \pi_j \times \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x^{(i)}-\mu_j)^T \Sigma^{-1} (x^{(i)}-\mu_j)\right)$$

$$= \Big(\pi_j \times \frac{1}{2} \Sigma^{-1}(x^{(i)}-\mu_K)(x^{(i)}-\mu_K)^T \Sigma^{-1} P(x^{(i)}|\mu_j,\Sigma)$$

$$+ \underbrace{\frac{\partial}{\partial \Sigma} |\Sigma|^{-1/2} (2\pi)^{-d/2}}_{-\frac{1}{2} \times (2\pi)^{-d/2} |\Sigma|^{-1/2} \Sigma^{-1}} \times \exp\left(-\frac{1}{2}(x^{(i)}-\mu_j)^T \Sigma^{-1} (x^{(i)}-\mu_j)\right)$$

ادامه صفحه بعد

$$\frac{\partial}{\partial \Sigma} \pi_j P(x^{(i)} \mid \mu_j, \Sigma) = \frac{1}{2} \pi_j \; P(x^{(i)} \mid \mu_j, \Sigma) \, \Sigma^{-1} \left(I + (x^{(i)} - \mu_K)(x^{(i)} - \mu_K)^T \Sigma^{-1} \right)$$

$$\Rightarrow \frac{\partial \text{ log-likelihood}}{\partial \Sigma} = \sum_{i=1}^{N} \frac{\sum_{j=1}^{K} \frac{1}{2} \pi_j P(x^{(i)} \mid \mu_j, \Sigma) \, \Sigma^{-1} \left(I + (x^{(i)} - \mu_K)(x^{(i)} - \mu_K)^T \Sigma^{-1} \right)}{\sum_{j'=1}^{K} \pi_{j'} P(x^{(i)} \mid \pi_{j'}, \Sigma)} = 0$$

$\Sigma \times \cdots \times 2 \Sigma$

$$\Rightarrow \sum_{i=1}^{N} \frac{\sum_{j=1}^{K} \pi_j P(x^{(i)} \mid \mu_j, \Sigma) \, \Sigma}{\sum_{j'=1}^{K} \pi_{j'} (P(x^{(i)} \mid \pi_{j'}, \Sigma)} =$$

$$= \sum_{i=1}^{N} \frac{\sum_{j=1}^{K} \pi_j P(x^{(i)} \mid \mu_j, \Sigma) (x^{(i)} - \mu_K)(x^{(i)} - \mu_K)^T}{\sum_{j'=1}^{K} \pi_{j'} (P(x^{(i)} \mid \pi_{j'}, \Sigma))}$$

$$\Rightarrow N \Sigma = \sum_{i=1}^{N} \frac{\sum_{j=1}^{N} \pi_j \; P(x^{(i)} \mid \mu_j, \Sigma) (x^{(i)} - \mu_K)(x^{(i)} - \mu_K)^T}{\sum_{j'=1}^{N} \pi_{j'} P(x^{(i)} \mid \pi_{j'}, \Sigma)}$$

$$\Rightarrow$$

$$\boxed{\Sigma = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{j=1}^{K} \pi_j P(x^{(i)} \mid \mu_j, \Sigma) (x^{(i)} - \mu_K)(x^{(i)} - \mu_K)^T}{\sum_{j'=1}^{K} \pi_{j'} P(x^{(i)} \mid \pi_{j'}, \Sigma)}}$$

سؤال دو :

۲.۱

E-step: $$\underset{Z|x,\theta}{E}[\log P(x,Z|\theta)] = \sum_{Z} \log P(x,Z|\theta)\, P(Z|x,\theta)$$

$$\Rightarrow P(Z_j^{(i)}=1 \mid x^{(i)},\theta) = \frac{P(x^{(i)}\mid Z_j^{(i)}=1,\theta)\, P(Z_j^{(i)}=1)}{\sum_{K=1}^{I} P(x^{(i)}\mid Z_K^{(i)}=1,\theta)\, P(Z_K^{(i)}=1)}$$

M_step: $$\theta^{t+1} = \underset{\theta}{\arg\max} \sum_{Z} P(Z|x,\theta^t) \log P(x,Z|\theta)$$

$$\frac{\partial}{\partial \theta_j} \sum_{i=1}^{N} \ln \sum_{j=1}^{I} \pi_j P_j(x^{(i)}) = \sum_{i=1}^{N} \ln \sum_{j=1}^{I} \pi_j \frac{e^{-\theta_j} \times \theta_j^{x^{(i)}}}{x^{(i)}!}$$

$$= \sum_{i=1}^{N} \frac{\frac{\partial}{\partial \theta_j} \pi_j \frac{e^{-\theta_j + x^{(i)}\ln\theta_j}}{x^{(i)}!}}{\sum_{K=1}^{I} \pi_K P_K(x^{(i)})} = \sum_{i=1}^{N} \frac{\pi_j \frac{e^{-\theta_j + x^{(i)}\ln\theta_j}}{x^{(i)}!} \times \left(-1 + \frac{x^{(i)}}{\theta_j}\right)}{\sum_{K=1} \pi_K P_K(x^{(i)})}$$

$$= \sum_{i=1}^{N} \frac{\pi_j P_j(x^{(i)}) \left(\frac{x^{(i)}}{\theta_j} - 1\right)}{\sum_{K=1} \pi_K P_K(x^{(i)})} = 0 \Rightarrow$$

$$\frac{1}{\theta_j} \sum_{i=1}^{N} \frac{\pi_j P_j(x^{(i)})\, x^{(i)}}{\sum_{K=1} \pi_K P_K(x^{(i)})} = \sum_{i=1}^{N} \frac{\pi_j P_j(x^{(i)})}{\sum_{K} \pi_K P_K(x^{(i)})} = N_K$$

$$\Rightarrow \theta_j^{new} = \frac{1}{N_K} \sum_{i=1}^{N} \frac{\pi_j P_j(x^{(i)})\, x^{(i)}}{\sum_{K=1} \pi_K P_K(x^{(i)})}$$ مانند حالت گوسی

π_K^{new} نیز می‌شود $\frac{N_K}{N}$ زیرا در سؤال قبل محاسبه شد و وابسته به توزیع نبود

$$\log P(x|\theta) = \log \sum_{z} P(x,z|\theta) = \log \sum_{z} \frac{Q(z)P(x,z|\theta)}{Q(z)}$$

نامساوی ینسن

$$\geq \sum_{z} Q(z)\left(\log P(x,z|\theta) - \log Q(z)\right)$$

این عبارت برای یک داده است و زمانی که داده ها مستقل باشند داریم:

$$\log P(X|\theta) = \sum_{x} \log P(x|\theta) \geq \sum_{x} \sum_{z} Q(z)\left(\log P(x,z,\theta) - \log Q(z)\right)$$

که ماکسیمم کردن $\ln P(X|\theta)$ به معنای ماکسیمم کردن تک تک $\log P(x|\theta)$ است یعنی:

$$\frac{\partial}{\partial Q(z)} \sum_{z} Q(z)\left(\log P(x,z|\theta) - \log Q(z)\right) + \lambda\left(\sum_{z} Q(z) - 1\right)$$

$$= \log P(x,z|\theta) - \log Q(z) - 1 + \lambda = 0$$

$$\Rightarrow \log Q(z) = \log P(x,z|\theta) + \lambda - 1$$

$$\Rightarrow Q(z) = P(x,z|\theta) \times C$$

$$\sum_{z'} Q(z') = 1 \rightarrow C \sum_{z'} P(x,z'|\theta) = C\, P(x|\theta) = 1$$

$$\Rightarrow C = \frac{1}{P(x|\theta)} = \frac{1}{\sum_{z'} P(x,z'|\theta)} \Rightarrow$$

$$\Rightarrow Q(z) = \frac{P(x,z|\theta)}{\sum_{z'} P(x,z'|\theta)} = \frac{P(x,z|\theta)}{P(x|\theta)} = P(z|x,\theta)$$

$$\Rightarrow Q(z) = P(z|x,\theta)$$

سؤال سه:

۳.۱

$$P(w_t, m_t \mid c_t) = \binom{m_t}{w_t} \times P_{c_t}^{w_t} (1-P_{c_t})^{m_t - w_t} \quad (II)$$

$$P(w_t, m_t \mid K) = \binom{m_t}{w_t} P_K^{w_t} (1-P_K)^{m_t - w_t} \quad (I)$$

۳.۲

E-Step: $$Q_t^{(i)}[K] = P(C_t = K \mid x_t, \theta^{i-1}) \propto$$

$$\propto P(x_t \mid C_t = K, \theta^{i-1}) \times P(C_t = K) =$$

$$= P(x_t \mid C_t = K, \theta^{i-1}) \times \Pi_K =$$

$$= \binom{m_t}{w_t} P_K^{w_t} (1-P_K)^{m_t - w_t} \Pi_K$$

$$\sum_{j=1}^{K} Q_t^{(i)}[j] = 1 \Rightarrow \boxed{Q_t^{(i)}[K] = \frac{\binom{m_t}{w_t} P_K^{w_t} (1-P_K)^{m_t - w_t} \Pi_K}{\sum_{j=1}^{K} \binom{m_t}{w_t} P_j^{w_t} (1-P_j)^{m_t - w_t} \Pi_j}}$$

M step: $$\underset{\theta}{\operatorname{argmax}} \underbrace{\sum_{t=1}^{n} \sum_{j=1}^{K} Q_t^{(i)}[j] \log P(x^{(t)} \mid P_j)}_{\ell}$$ ۳.۳

$$\frac{\partial \ell}{\partial P_K} = \sum_{t=1}^{n} Q_t^{(i)}[K] \frac{\partial}{\partial P_K} \log \binom{m_t}{w_t} P_K^{w_t} (1-P_K)^{m_t - w_t},$$

$$\frac{\partial}{\partial P_K} P_K^{w_t} = \frac{w_t}{P_K}$$

$$\frac{\partial (1-P_K)^{m_t - w_t}}{\partial P_K} = \frac{-(m_t - w_t)}{1-P_K}$$

$$\Rightarrow \sum_{t=1}^{n} Q_t^{(i)}[K] \left(\frac{w_t}{P_K} - \frac{(m_t - w_t)}{1-P_K} \right) = 0$$

$$\Rightarrow \frac{\sum_{t=1}^{n} Q_t^{(i)}[K] w_t}{P_K} = \frac{\sum_{t=1}^{n} Q_t^{(i)} (m_t - w_t)}{1-P_K} \Rightarrow \boxed{P_K = \frac{\sum_{t=1}^{n} Q_t^{(i)}[K] w_t}{\sum_{t=1}^{n} Q_t^{(i)}[K] m_t}}$$

Scanned by CamScanner

سوال چهار :

۴.۱

$$P(\theta|X) \propto P(X|\theta)P(\theta)$$

① M-Step :

$$\underset{\theta}{\operatorname{argmax}} \ln P(X|\theta)P(\theta) = \sum_{z} P(z|X,\theta^{old}) \ln P(X,z|\theta)P(\theta)$$

$$= \sum_{z} P(z|X,\theta^{old})(\ln P(X,z|\theta) + \ln P(\theta))$$

$$= \underbrace{\sum_{z} P(z|x,\theta^{old}) \ln P(x,z|\theta)}_{\text{مانند قبل } Q} + \underbrace{\sum_{z} P(z|x,\theta^{old})}_{1} \ln P(\theta)$$

$$= Q + \ln P(\theta)$$

پس در M-Step در این حالت فقط یک $\ln P(\theta)$ اضافه می‌شود

② E-step : مانند حالت قبل است.

$$E_{z|x,\theta}[\log P(x,z|\theta)] = \sum_{z} \log P(x,z|\theta)\, P(z|x,\theta^{old})$$

۴.۲ برچسب دسته داده : Z

E step :
$$P(A \mid X = A \text{ یا } C) = \frac{1}{1+2\theta}$$

$$P(C \mid X = A \text{ یا } C) = \frac{2\theta}{1+2\theta}$$

در بقیه حالات این احتمال یا یک یا صفر است زیرا برچسب داده مشخص است.

M-Step:

$$\theta^{t+1} = \underset{\theta}{\operatorname{argmax}}\ Q + \ln P(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{4} P(z|x,\theta^{t}) \ln P(x,z|\theta) + \ln P(\theta)$$

$$= \underset{\theta}{\operatorname{argmax}}\ n_b \times \ln\left(\tfrac{1}{3}(1-\theta)\right) + n_d \times \ln\left(\tfrac{1}{3}(1-\theta)\right) +$$

$$(n_a+n_c)\left(\frac{1}{1+2\theta^t} \ln \frac{1}{3} + \frac{2\theta^t}{1+2\theta^t} \ln\left(\tfrac{2}{3}\theta\right)\right) + n\left((v_1-1)\ln\theta + (v_2-1)\ln(1-\theta)\right)$$

$$\Rightarrow \frac{\partial}{\partial\theta} Q + \ln P(\theta)$$

ادامه محاسبات در صفحه بعد

$$\frac{\partial}{\partial \theta} Q + \ln P(\theta) = -nb\left(\frac{1}{1-\theta}\right) + -nd\left(\frac{1}{1-\theta}\right) +$$

$$+ (na + nc) \times \left(\frac{2\theta^{t}}{1+2\theta^{t}} \times \frac{1}{\theta}\right) + \frac{n(v_1 - 1)}{\theta} - \frac{n(v_2 - 1)}{1-\theta} = 0$$

$$\Rightarrow \frac{(nb + nd) + n(v_2 - 1)}{1-\theta} = \left(\frac{2\theta^{t}(na + nc)}{(1+2\theta^{t})} + n(v_1 - 1)\right) \times \frac{1}{\theta}$$

$$\Longrightarrow \theta = \frac{\frac{2\theta^{t}}{1+2\theta^{t}}(na + nc) + n(v_1 - 1)}{\frac{2\theta^{t}}{1+2\theta^{t}}(na + nc) + n(v_1 - 1) + n(v_2 - 1) + nb + nd}$$