سوال ۳

$$D = \{x_1, \dots, x_n\}$$

$$L(\theta) = \log P(D|\theta) = \log P(x_1, \dots, x_n|\theta) \overset{i.i.d.}{=\!=} \log \prod_{i=1}^{n} P(x_i|\theta) = \sum_{i=1}^{n} \log P(x_i|\theta)$$

$$= \sum_{i=1}^{n} \log 2x_i - 2\log\theta$$

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^{n} \frac{-2}{\theta} = \frac{-2n}{\theta} < 0$$

با توجه به اینکه مشتق نسبت به θ همواره منفی است، لذا تابع likelihood نسبت به θ اکیداً نزولی است. بنابراین برای بیشینه کردن آن باید کوچکترین مقدار ممکن برای θ را انتخاب کنیم. از آن جایی که تمام x_i ها در رابطه‌ی $0 \leq x_i \leq \theta$ صدق می‌کنند، لذا باید θ از همه‌ی x_i ها بزرگتر باشد. پس کوچکترین مقدار ممکن برای θ به صورت زیر است:

$$\hat{\theta}_{ML} = \max\{x_1, \dots, x_n\} = x_{max}$$

سوال ۶

فرض کنیم داده های $\{x_1, \dots, x_n\}$ را در اختیار داریم. برای هر یک از داده ها، یک متغیر نهان z_i در نظر می‌گیریم. در واقع متغیر z_i مشخص می‌کند که نمونه x_i از کدام مؤلفه گاوسی آمده است. بنابراین z_i یک بردار K-بعدی به صورت one-hot خواهد بود که فقط یکی از درایه‌های آن 1 و مابقی صفر هستند. توزیع احتمال توأم x_i و z_i به صورت زیر خواهد بود:

$$P(x_i, z_i) = \prod_{j=1}^{K} \left(p_j \, \mathcal{N}(x_i ; \mu_j , \Sigma_j) \right)^{z_{ij}}$$

در رابطه‌ی بالا K تعداد مؤلفه ها، z_{ij} درایه‌ی j-ام بردار z_i، p_j وزن (احتمال) مولفه j-ام و μ_j و Σ_j به ترتیب میانگین و ماتریس کوواریانس مولفه j-ام هستند.

$$\log P(x_1, \ldots, x_n, z_1, \ldots, z_n \mid \theta) = \log \prod_{i=1}^{n} P(x_i, z_i \mid \theta) = \sum_{i=1}^{n} \log P(x_i, z_i \mid \theta)$$

$$= \sum_{i=1}^{n} \log \prod_{j=1}^{K} \left(p_j \, \mathcal{N}(x_i ; \mu_j, \Sigma_j)\right)^{z_{ij}} = \sum_{i=1}^{n} \sum_{j=1}^{K} z_{ij} \left(\log p_j + \log \mathcal{N}(x_i ; \mu_j, \Sigma_j)\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij} \left(\log p_j - \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)\right)$$

$$Q(\theta, \theta^t) = E\left[\log P(x_1, \ldots, x_n, z_1, \ldots, z_n \mid \theta)\right] =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} E[z_{ij}] \left(\log p_j - \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^T (x_i - \mu_j)\right)$$

حال باید از $Q(\theta,\theta^t)$ نسبت به پارامترها مشتق بگیریم:

$$\frac{\partial Q}{\partial \mu_j} = \sum_{i=1}^{n} E[z_{ij}]\, \Sigma_j^{-1} (x_i-\mu_j) = 0 \xrightarrow[\text{ضرب می‌کنیم } \Sigma_j \text{ در}]{\text{طرفین را از چپ}} \sum_{i=1}^{n} E[z_{ij}]\,(x_i-\mu_j) = 0$$

$$\Rightarrow \boxed{\mu_j = \frac{\sum_{i=1}^{n} E[z_{ij}]\, x_i}{\sum_{i=1}^{n} E[z_{ij}]}}$$

$$\frac{\partial Q}{\partial \Sigma_j} = \sum_{i=1}^{n} \left(-\frac{1}{2}\Sigma_j^{-1} + \frac{1}{2}\Sigma_j^{-1}(x_i-\mu_j)(x_i-\mu_j)^T\Sigma_j^{-1}\right) E[z_{ij}] = 0$$

طرفین را از چپ و راست در Σ_j ضرب می‌کنیم

$$\Rightarrow \sum_{i=1}^{n}\left(-\frac{1}{2}\Sigma_j + \frac{1}{2}(x_i-\mu_j)(x_i-\mu_j)^T\right)E[z_{ij}] = 0 \Rightarrow \boxed{\Sigma_j = \frac{\sum_{i=1}^{n} E[z_{ij}](x_i-\mu_j)(x_i-\mu_j)^T}{\sum_{i=1}^{n} E[z_{ij}]}}$$

برای بهینه سازی نسبت به p_j ها باید به قید $\sum_{j=1}^{K} p_j = 1$ توجه کنیم و یک ترم لاگرانژ به صورت $\lambda\left(\sum_{j=1}^{K} p_j - 1\right)$

به Q اضافه کنیم و سپس نسبت به p_j مشتق بگیریم:

$$\frac{\partial}{\partial p_j}\left(Q - \lambda\left(\sum_{j=1}^{K} p_j - 1\right)\right) = \sum_{i=1}^{n} E[z_{ij}]\left(\frac{1}{p_j}\right) - \lambda = 0$$

$$\Longrightarrow \quad p_j = \frac{\sum_{i=1}^{n} E[z_{ij}]}{\lambda} \qquad j = 1, \dots, K$$

$$\sum_{j=1}^{K} p_j = \sum_{j=1}^{K} \frac{\sum_{i=1}^{n} E[z_{ij}]}{\lambda} = 1 \Longrightarrow \lambda = \sum_{j=1}^{K}\sum_{i=1}^{n} E[z_{ij}] = n \Longrightarrow \boxed{p_j = \frac{\sum_{i=1}^{n} E[z_{ij}]}{n}}$$