

CS 7641 CSE/ISYE 6740 Homework 3

Tong Yiming

1 EM application

1.(a)

Note that $x^{(pr)} = y^{(pr)} + z^{(pr)} + \epsilon^{(pr)}$, where $\epsilon^{(pr)} \sim \mathcal{N}(0, \sigma^2)$, and $y^{(pr)}, z^{(pr)}, \epsilon^{(pr)}$ are independent with each other. We could therefore obtain

$$\begin{aligned} E(x^{(pr)}) &= E(y^{(pr)} + z^{(pr)} + \epsilon^{(pr)}) = \mu_p + \nu_r, \\ Cov(x^{(pr)}, y^{(pr)}) &= Cov(y^{(pr)} + z^{(pr)} + \epsilon^{(pr)}, y^{(pr)}) = \sigma_p^2, \\ Cov(x^{(pr)}, z^{(pr)}) &= \tau_r^2, \\ \sigma_{x^{(pr)}}^2 &= \sigma_p^2 + \tau_r^2 + \sigma^2. \end{aligned}$$

Thus, the mean vector and covariance matrix read

$$\mu = \begin{pmatrix} \mu_p \\ \nu_r \\ \mu_p + \nu_r \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_p^2 & 0 & \sigma_p^2 \\ 0 & \tau_r^2 & \tau_r^2 \\ \sigma_p^2 & \tau_r^2 & \sigma_p^2 + \tau_r^2 + \sigma^2 \end{pmatrix}.$$

1.(b)

The desired set of parameters to be estimated is $\theta = \{\mu_p, \sigma_p^2, \nu_r, \tau_r^2\}$. From the conditional distribution formula of jointly Gaussian distribution, we have

$$\begin{aligned} (y^{(pr)}, z^{(pr)})|x^{(pr)} &\sim \mathcal{N}(\mu', \Sigma'), \text{ where} \\ \mu' &= \frac{1}{\sigma_p^2 + \tau_r^2 + \sigma^2} \begin{pmatrix} \mu_p(\tau_r^2 + \sigma^2) + \sigma_p^2(x^{(pr)} - \nu_r) \\ \nu_r(\sigma_p^2 + \sigma^2) + \tau_r^2(x^{(pr)} - \mu_p) \end{pmatrix}, \\ \Sigma' &= \frac{1}{\sigma_p^2 + \tau_r^2 + \sigma^2} \begin{pmatrix} \sigma_p^2(\tau_r^2 + \sigma^2) & -\sigma_p^2\tau_r^2 \\ -\sigma_p^2\tau_r^2 & \tau_r^2(\sigma_p^2 + \sigma^2) \end{pmatrix}. \end{aligned}$$

Thus,

$$Q^{(pr)}(\theta'|\theta) = \iint_{y^{(pr)}, z^{(pr)}} p(y^{(pr)}, z^{(pr)}|x^{(pr)}; \theta) \log p(y^{(pr)}, z^{(pr)}, x^{(pr)}; \theta') dy^{(pr)} dz^{(pr)},$$

where

$$\begin{aligned}
p(y^{(pr)}, z^{(pr)} | x^{(pr)}) &= \text{PDF}(\mathcal{N}(y^{(pr)}, z^{(pr)} | x^{(pr)})) \\
&= \frac{1}{2\pi\sqrt{\det \Sigma'}} \exp\left(-\frac{1}{2}(x - \mu')^T \Sigma'^{-1}(x - \mu')\right), \\
p(y^{(pr)}, z^{(pr)}, x^{(pr)}) &= \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right).
\end{aligned}$$

2.

The optimization problem in M-step is

$$\begin{aligned}
\theta_{\text{new}} &= \arg \max_{\theta'} \sum_{p,r} Q^{(pr)}(\theta' | \theta) \\
&= \arg \max_{\theta'} \sum_{p,r} \iint_{y^{(pr)}, z^{(pr)}} p(y^{(pr)}, z^{(pr)} | x^{(pr)}; \theta) \log p(y^{(pr)}, z^{(pr)}, x^{(pr)}; \theta') dy^{(pr)} dz^{(pr)}. \\
&= \arg \max_{\theta'} \sum_{p,r} \mathbb{E}[\text{PDF}(\mathcal{N}(y^{(pr)} + z^{(pr)}, \sigma)) \cdot \text{PDF}(\mathcal{N}(\mu_p, \sigma_p)) \cdot \text{PDF}(\mathcal{N}(\nu_r, \tau_r))] \\
&= \arg \max_{\theta'} \sum_{p,r} \mathbb{E}\left[\log \frac{1}{\sigma_p \tau_r} - \frac{1}{2\sigma_p^2} (y^{(pr)} - \mu_p)^2 - \frac{1}{2\tau_r^2} (z^{(pr)} - \nu_r)^2\right].
\end{aligned}$$

Here we could take the expectation of $y^{(pr)}$ and $z^{(pr)}$, which follows the conditional distribution we derived before

$$\begin{aligned}
\mathbb{E}(y^{(pr)}) &= \mu'_1, \mathbb{E}(z^{(pr)}) = \mu'_2, \\
\mathbb{E}[(y^{(pr)})^2] &= \mu_1'^2 + \Sigma'_{11}, \mathbb{E}[(z^{(pr)})^2] = \mu_2'^2 + \Sigma'_{22},
\end{aligned}$$

where the subscripts denote the position of an matrix of array element. Finally we come to an analytical form of the target function,

$$\theta_{\text{new}} = \arg \max_{\theta'} \sum_{p,r} \log \frac{1}{\sigma_p \tau_r} - \frac{1}{2\sigma_p^2} (\Sigma'_{11} + \mu_1'^2 - 2\mu'_1 \mu_p + \mu_p^2) - \frac{1}{2\tau_r^2} (\Sigma'_{22} + \mu_2'^2 - 2\mu'_2 \nu_r + \nu_r^2).$$

With some derivative taken, we have the updated parameters as

$$\begin{aligned}
\mu_p &= \frac{1}{R} \sum_r \mu'_1, \\
\nu_r &= \frac{1}{P} \sum_p \mu'_2, \\
\sigma_p^2 &= \frac{1}{R} \sum_r \Sigma'_{11} + \mu_1'^2 - 2\mu_p \mu_1'^2 + \mu_p^2, \\
\tau_r^2 &= \frac{1}{P} \sum_p \Sigma'_{22} + \mu_2'^2 - 2\tau_r \mu_2'^2 + \nu_r^2.
\end{aligned}$$

2 Basic optimization

1.

We first calculate the Hessian:

$$\begin{aligned}
 \nabla f &= \frac{\sum_i e^{x_i} \hat{i}}{\sum_i e^{x_i}}, \\
 \nabla^2 f &= \nabla(\nabla f) = \nabla \frac{1}{\sum_i e^{x_i}} \sum_i e^{x_i} \hat{i} + \frac{1}{\sum_i e^{x_i}} \nabla^2 \sum_i e^{x_i} \\
 &= -\frac{\sum_i e^{x_i} \hat{i}}{(\sum_i e^{x_i})^2} \sum_i e^{x_i} \hat{i} + \frac{1}{\sum_i e^{x_i}} \sum_i e^{x_i} \hat{i} \hat{i} \\
 &= \frac{1}{(\sum_i e^{x_i})^2} \sum_{i,j} (\delta_{i,j} e^{x_i} \sum_i e^{x_i} - e^{x_i} e^{x_j}) \hat{i} \hat{j}
 \end{aligned}$$

Next, we show that the $\nabla^2 f \succeq 0$. Take v as any given vector, we have

$$\begin{aligned}
 v^T \cdot \nabla^2 f \cdot v &= \left(\sum_i v_i \hat{i} \right) \cdot \frac{1}{(\sum_i e^{x_i})^2} \sum_{i,j} (\delta_{i,j} e^{x_i} \sum_i e^{x_i} - e^{x_i} e^{x_j}) \hat{i} \hat{j} \cdot \left(\sum_i v_i \hat{i} \right) \\
 &= \frac{1}{(\sum_i e^{x_i})^2} \left[\sum_i e^{x_i} \sum_i e^{x_i} v_i^2 - \left(\sum_i e^{x_i} v_i \right)^2 \right] \geq 0,
 \end{aligned}$$

which is Cauchy-Schwarz inequality.

2.

True for $X \succ 0$. We verify its concavity by considering the function $f(t) = \log \det(X + tV)$ for any direction V . We only take consider a neighbourhood of X where $X + tV \succ 0$. We have

$$\begin{aligned}
 f(t) &= \log \det(X + tV) = \log[\det X^{1/2} \det(I + tX^{-1/2} V X^{-1/2}) \det X^{1/2}] \\
 &= \log \det X + \log \det(I + tX^{-1/2} V X^{-1/2}) = \sum_i \log(1 + t\lambda_i) + \log \det X,
 \end{aligned}$$

where λ_i is the i^{th} eigenvalue of $X^{-1/2} V X^{-1/2}$. Now it's easy to calculate

$$f''(t) = -\sigma_i \frac{\lambda_i}{(1 + t\lambda_i)^2} < 0,$$

which infers $f(t)$ is concave. Thus $\log \det X$ is concave.

3.

We show the result by induction on m .

(1) basis

$m = 1$ is trivial.

(2) inductive step

if for case $m \leq k$ the proposition holds, then for $m = k + 1$:

$$\begin{aligned} f\left(\sum_{i=1}^{k+1} \alpha_i x_i\right) &= f\left(\alpha_1 x_1 + \sum_{i=2}^{k+1} \alpha_i x_i\right) \geq \alpha_1 f(x_1) + (1 - \alpha_1) f\left(\sum_{i=2}^{k+1} \frac{\alpha_i}{1 - \alpha_1} x_i\right) \\ &\geq \alpha_1 f(x_1) + (1 - \alpha_1) \sum_{i=2}^{k+1} \frac{\alpha_i}{1 - \alpha_1} x_i = \sum_{i=1}^{k+1} \alpha_i x_i. \end{aligned}$$

(3) conclusion

The result(Jensen inequality) holds for all natural numbers.

3 Solve simple optimization problems

1)

It's easy to verify that V_p is a convex set. Since $\nabla f = c$, the minimum could only be on the boundaries. From KKT conditions we know that $\nabla f + \mu \nabla g = 0$, where $\mu > 0$. (The constraint should be considered for each quadrant. But we know that the minimizer is either in one quadrant or on an axis. Therefore we only need to take one quadrant in to equation and deal with the corners, which could be easily proven not the minimizer) Thus,

$$\mu \nabla g = \mu \sum_i \text{sgn}(x_i) p(\text{sgn}(x_i) x_i)^{p-1} \hat{i} = -c,$$

where sgn is the sign function. Therefore we have

$$\begin{aligned} \text{sgn}(x_i) &= -\text{sgn}(c_i), \\ x_i &= -\text{sgn}(c_i) \left(\frac{|c_i|}{\mu p}\right)^{\frac{1}{p-1}}, \end{aligned}$$

where c_i is the i^{th} component of c . Finally from $\sum_i |x_i|^p = 1$ we could normalize μ as $\mu = \left(\sum_i \left(\frac{|c_i|}{p}\right)^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}}$

2)

Added Lagrange multiplier, the target function becomes

$$\mathcal{L} = d^T x + \sum_i x_i \log x_i + \lambda \left(\sum_i x_i - 1\right).$$

Thus we have

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_i} &= d_i + \lambda + 1 + \log x_i = 0, \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= \sum_i x_i - 1 = 0. \end{aligned}$$

Solving these equations, we could obtain the minimizer as $x_i = e^{-(d_i + \lambda + 1)}$, where $\lambda = -\log(\sum_i e^{-(d_i + 1)})$.