Homework 3 - Part 1 - Math 534

Elijah Amirianfar

2024-02-13

Exercise J-2.2: In this exercise, we assume that we have a set of data generated from a p-variate normal with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_p)^T$ and a $p \times p$ covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$

Let the following log-likelihood function be:

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n}) = -\frac{1}{2}\left(np\log(2\pi) + n\log|\boldsymbol{\Sigma}| + trace(\boldsymbol{\Sigma^{-1}}C(\boldsymbol{\mu}))\right)$$

Where $c(\boldsymbol{\mu}) =$

$$\sum_{i=1}^{n}(\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^T$$

Problem 1: Obtain formulas for the elements of the gradient and hessian of $\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Computing Gradients/1st Derivatives

Here we will compute the following derivatives:

$$\frac{\partial \ell}{\partial \mu_i}, \ \frac{\partial \ell}{\partial \sigma_{ii}}, \ \frac{\partial \ell}{\partial \sigma_{ij}}$$

Thus, we start with our Log Likelihood.

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n}) = -\frac{1}{2}\left(np\log(2\pi) + n\log|\boldsymbol{\Sigma}| + trace(\boldsymbol{\Sigma^{-1}}\sum_{i=1}^{n}[(\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T]))\right)$$

$$\partial \ell(\partial \boldsymbol{\mu}) = \frac{1}{2} trace\left(\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}[\partial \boldsymbol{\mu}(\boldsymbol{x}_i - \boldsymbol{\mu})^T + (\boldsymbol{x}_i - \boldsymbol{\mu})\partial \boldsymbol{\mu}^T]\right)$$

Since the following is true:

$$trace(\boldsymbol{\Sigma}^{-1}\partial \boldsymbol{\mu}(\boldsymbol{x}_i - \boldsymbol{\mu})^T) = trace(\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\partial \boldsymbol{\mu}^T),$$

We can obtain the following:

$$= trace\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})\partial \boldsymbol{\mu}^T\right)$$

Thus, we get the final value:

$$\frac{\partial \ell}{\partial \mu_i} = \left[\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})\right]_i$$

Now, we compute the following two derivatives: $\dfrac{\partial \ell}{\partial \sigma_{ii}}$, $\dfrac{\partial \ell}{\partial \sigma_{ii}}$

$$
\begin{aligned}
\partial \ell(\partial \boldsymbol{\Sigma}) &= -\frac{n}{2} trace(\boldsymbol{\Sigma}^{-1}\partial \boldsymbol{\Sigma}) - \frac{1}{2} trace(\boldsymbol{\Sigma}^{-1}\partial \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T) \\
&= -\frac{n}{2} trace\left(\boldsymbol{\Sigma}^{-1}\partial \boldsymbol{\Sigma} + \frac{1}{n}\Sigma^{-1}\partial \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\right) \\
&= -\frac{n}{2} trace\left(\boldsymbol{\Sigma}^{-1}\partial \boldsymbol{\Sigma} + \frac{1}{n}\Sigma^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}\partial \boldsymbol{\Sigma}\right) \\
&= -\frac{n}{2} trace\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\right)\boldsymbol{\Sigma}^{-1}\partial \boldsymbol{\Sigma}\right)
\end{aligned}
$$

Let $\boldsymbol{A} = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\right)\boldsymbol{\Sigma}^{-1}$

Thus,

$$
\partial \ell(\partial \boldsymbol{\Sigma}) = -\frac{n}{2} trace(\boldsymbol{A}\partial \boldsymbol{\Sigma})
$$

Which implies that

$$
\begin{aligned}
\frac{\partial \ell}{\partial \sigma_{ii}} &= -\frac{n}{2} A_{ii} \\
\frac{\partial \ell}{\partial \sigma_{ij}} &= -\frac{n}{2}[A_{ij} + A_{ji}]
\end{aligned}
$$

Computing Hessians/2nd Derivatives

Here we will compute the following derivatives:

$$\frac{\partial^2 \ell}{\partial \mu_i \partial \mu_i},\ \frac{\partial^2 \ell}{\partial \mu_i \partial \mu_j},\ \frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \mu_k},\ \frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \mu_k},\ \frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \sigma_{kk}},\ \frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}},\ \frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kk}},\ \frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \sigma_{kl}}$$

From the previous part, we found that

$$\begin{aligned}
\partial \ell(\partial \boldsymbol{\mu}) &= trace\left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu}) \partial \boldsymbol{\mu}^T\right) \\
\partial \ell(\partial \boldsymbol{\Sigma}) &= -\frac{n}{2} trace\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^T\right) \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma}\right)
\end{aligned}$$

Now we need to compute $\partial\partial \ell(\partial \boldsymbol{\mu}, \partial \boldsymbol{\mu})$, $\partial\partial \ell(\partial \boldsymbol{\Sigma}, \partial \boldsymbol{\Sigma})$, and $\partial\partial \ell(\partial \boldsymbol{\mu}, \partial \boldsymbol{\Sigma})$

Thus, we get:

$$\begin{aligned}
\partial\partial \ell(\partial \boldsymbol{\mu}, \partial \boldsymbol{\mu}) &= trace\left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} -\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T\right) \\
&= -n\ trace\left(\boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T\right) \\
\partial\partial \ell(\partial \boldsymbol{\mu}, \partial \boldsymbol{\Sigma}) &= -trace\left(\boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu}) \partial \boldsymbol{\mu}^T\right) = \partial\partial \ell(\partial \boldsymbol{\Sigma}, \partial \boldsymbol{\mu}) \\
\partial\partial \ell(\partial \boldsymbol{\Sigma}, \partial \boldsymbol{\Sigma}) &= -\frac{1}{2} trace\{-\boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1}\left(n\boldsymbol{\Sigma} - \sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\right) \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \\
&+ \boldsymbol{\Sigma}^{-1}(n \partial \boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \\
&+ \boldsymbol{\Sigma}^{-1}\left(n\boldsymbol{\Sigma} - \sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\right)\left(-\boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1}\right) \partial \boldsymbol{\Sigma}\} \\
&= trace\left\{\boldsymbol{\Sigma}^{-1}\left(n\boldsymbol{\Sigma} - \sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\right) \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} - \frac{n}{2} \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma}\right\} \\
&= ntrace\left\{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\right) \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} - \frac{1}{2} \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma}\right\} \\
&= ntrace\left\{\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\right) - \frac{1}{2} I\right] \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \partial \boldsymbol{\Sigma}\right\}
\end{aligned}$$

To compute $\frac{\partial^2 \ell}{\partial \mu_i \partial \mu_i}$, $\frac{\partial^2 \ell}{\partial \mu_i \partial \mu_j}$, we know that $\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T$ is a $p \times p$ matrix where it is zero everywhere except at the ii'th or the ij'th element. Thus, we get the following:

$$\frac{\partial^2 \ell}{\partial \mu_i \partial \mu_i} = [-n\boldsymbol{\Sigma}^{-1}]_{ii}, \quad \frac{\partial^2 \ell}{\partial \mu_i \partial \mu_j} = [-n\boldsymbol{\Sigma}^{-1}]_{ij}$$

To compute $\frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \mu_k}$, $\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \mu_k}$, we let A $= \boldsymbol{\Sigma}^{-1}$ and $p =$ number of dimensions present, we get the following:

$$\frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \mu_k} = -\sum_{w=1}^{p} a_{iw} a_{ki} \left[\sum_{z=1}^{n} (\boldsymbol{x}_z - \boldsymbol{\mu}) \right]_w, \quad \frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \mu_k} = -\sum_{w=1}^{p} (a_{iw} a_{kj} + a_{jw} a_{ki}) \left[\sum_{z=1}^{n} (\boldsymbol{x}_z - \boldsymbol{\mu}) \right]_w$$

To compute $\frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \sigma_{kk}}$, $\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}}$, $\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kk}}$, $\frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \sigma_{kl}}$, we let :

$$A = \left[\boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\Sigma} - \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T \right) - \frac{1}{2} I \right] \boldsymbol{\Sigma}^{-1}, \ B = \boldsymbol{\Sigma}^{-1}$$

Thus, we end up with the following differential: $ntrace(A \partial \boldsymbol{\Sigma} B \partial \boldsymbol{\Sigma})$

$$\frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \sigma_{kk}} = n A_{ki} B_{ik}$$

$$\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}} = n(A_{kj} B_{il} + A_{ki} B_{jl} + A_{lj} B_{ik} + A_{li} B_{jk})$$

$$\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kk}} = n(A_{kj} B_{ik} + A_{ki} B_{jk})$$

$$\frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \sigma_{kl}} = n(A_{ki} B_{il} + A_{li} B_{ik})$$

Problem 2: Obtain formulas for the elements of the information matrix.

$$-E\left[\frac{\partial^2 \ell}{\partial \mu_i \partial \mu_j}\right] = [n\boldsymbol{\Sigma}^{-1}]_{ij} \text{ (For both cases, this holds)}$$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \mu_k}\right] = 0 \text{ (For all cases, they will be zero because they are multiplied by } E(\boldsymbol{x}_i - \boldsymbol{\mu}) = 0)$$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \sigma_{kk}}\right] = \frac{n}{2} \boldsymbol{\Sigma}_{ki}^{-1} \boldsymbol{\Sigma}_{ik}^{-1}$$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}}\right] = \frac{n}{2} (\boldsymbol{\Sigma}_{kj}^{-1} \boldsymbol{\Sigma}_{il}^{-1} + \boldsymbol{\Sigma}_{ki}^{-1} \boldsymbol{\Sigma}_{jl}^{-1} + \boldsymbol{\Sigma}_{lj}^{-1} \boldsymbol{\Sigma}_{ik}^{-1} + \boldsymbol{\Sigma}_{li}^{-1} \boldsymbol{\Sigma}_{jk}^{-1})$$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kk}}\right] = \frac{n}{2} (\boldsymbol{\Sigma}_{kj}^{-1} \boldsymbol{\Sigma}_{ik}^{-1} + \boldsymbol{\Sigma}_{ki}^{-1} \boldsymbol{\Sigma}_{jk}^{-1})$$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma_{ii} \partial \sigma_{kl}}\right] = \frac{n}{2} (\boldsymbol{\Sigma}_{ki}^{-1} \boldsymbol{\Sigma}_{il}^{-1} + \boldsymbol{\Sigma}_{li}^{-1} \boldsymbol{\Sigma}_{ik}^{-1})$$