Homework 2, STAT 5205

Zongyi Liu

Mon, Feb 17, 2025

Question 1

Consider the Bayesian model: G is a categorical random variable which takes values in $\{1,2\}$ and satisfies

$$\mathbb{P}(G=1)=\pi_1, \quad \mathbb{P}(G=2)=\pi_2$$

with $\pi_1+\pi_2=1$. $\boldsymbol{X}$ is a d-dimensional random vector $\left(\boldsymbol{X}\in\mathbb{R}^d\right)$ with conditional Gaussian distributions

$$\boldsymbol{X}\,|G=1\sim N\left(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}\right) \quad \boldsymbol{X}|\,G=2\sim N\left(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}\right)$$

Compute the following items (write out the formulas in terms of $\pi_1,\pi_2,\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\boldsymbol{\Sigma}$).

Part i

The joint distributions $P\{G=1,\boldsymbol{X}=\boldsymbol{x}\}=?$, $P(G=2,\boldsymbol{X}=\boldsymbol{x})=?$

By the law of conditional probability:

$$P(G=g,X=x)=P(G=g)P(X=x|G=g)$$

For $G=1$:

$$P(G=1,X=x)=\pi_1\cdot\frac{1}{(2\pi)^{d/2}(det(\Sigma))^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)$$

For $G=2$:

$$P(G=2,X=x)=\pi_2\cdot\frac{1}{(2\pi)^{d/2}(det(\Sigma))^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right)$$

Part ii

The marginal distributional density of $\boldsymbol{X}, d(\boldsymbol{x})=?$

The marginal density of X, denoted as $d(x)$, can be found by summing the joint densities over all possible values of G:

$$d(x)=P(X=x)=P(G=1,X=x)+P(G=2,X=x)$$

Using the joint distributions derived earlier:

$$d(x)=\pi_1 f_1(x)+\pi_2 f_2(x)$$

Thus the marginal density of X is:

$$d(x)=\pi_1\frac{1}{(2\pi)^{d/2}(det(\Sigma))^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)+\pi_2\frac{1}{(2\pi)^{d/2}(det(\Sigma))^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right)$$

Part iii

The conditional distribution $P\{G=1\mid\boldsymbol{X}=\boldsymbol{x}\}=?$

We can solve this problem with the help of Baye's Rule

$$P(G=1\mid X=x)=\frac{P(G=1,X=x)}{P(X=x)}$$

From our previous derivations:

The joint distribution:

$$P(G=1,X=x)=\pi_1 f_1(x)$$

where

$$f_1(x) = \frac{1}{(2\pi)^{d/2}(det(\Sigma))^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)\right)$$

And the marginal density of X:

$$P(X = x) = d(x) = \pi_1 f_1(x) + \pi_2 f_2(x)$$

where

$$f_2(x) = \frac{1}{(2\pi)^{d/2}(det(\Sigma))^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right)$$

So the conditional distribution is

$$P(G = 1 \mid X = x) = \frac{\pi_1 f_1(x)}{\pi_1 f_1(x) + \pi_2 f_2(x)}$$

Question 2

$f(\boldsymbol{x})$ is a function of four variables/arguments, and has the formula below

$$f\left(x_1, x_2, x_3, x_4\right) = x_4 e^{-\left(x_1-1\right)^2/2} + \log\left(x_2^2+1\right) + \cos\left(x_3\right)\sin\left(x_4^2\right) + \frac{\sqrt{x_2}}{x_3^2+2}$$

Compute the first and second order gradients of $f(\boldsymbol{x})$, i.e.,

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = ?$$
$$\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x}\partial \boldsymbol{x}^\top} = ?$$

Part i For the first order problem, we have to compute
$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}, \frac{\partial f}{\partial x_4}\right)$.
Partial derivative with respect to x_1:

$$\frac{\partial f}{\partial x_1} = x_4 \cdot e^{-\frac{(x_1-1)^2}{2}} \cdot \left(-\frac{x_1-1}{1}\right) = -x_4(x_1-1)e^{-\frac{(x_1-1)^2}{2}}$$

Partial derivative w.r.t. x_2:

$$\frac{\partial f}{\partial x_2} = \frac{2x_2}{x_2^2+1} + \frac{1}{2(x_3^2+2)\sqrt{x_2}}$$

Partial derivative w.r.t. x_3:

$$\frac{\partial f}{\partial x_3} = -\sin(x_3)\sin(x_4^2) - \frac{2x_3\sqrt{x_2}}{(x_3^2+2)^2}$$

Partial derivative w.r.t. x_4:

$$\frac{\partial f}{\partial x_4} = e^{-\frac{(x_1-1)^2}{2}} + \cos(x_3)\cos(x_4^2)\cdot 2x_4$$

Thus the gradient vector is:

$$\nabla f = \begin{bmatrix} -x_4(x_1-1)e^{-\frac{(x_1-1)^2}{2}} \\ \frac{2x_2}{x_2^2+1} + \frac{1}{2(x_3^2+2)\sqrt{x_2}} \\ -\sin(x_3)\sin(x_4^2) - \frac{2x_3\sqrt{x_2}}{(x_3^2+2)^2} \\ e^{-\frac{(x_1-1)^2}{2}} + \cos(x_3)\cos(x_4^2)\cdot 2x_4 \end{bmatrix}$$

Part ii The Hessian matrix H_f is a 4×4 symmetric matrix whose entries are the second-order partial derivatives of $f(x)$:

$$\boxed{H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\partial x_2} & \frac{\partial^2 f}{\partial x_1\partial x_3} & \frac{\partial^2 f}{\partial x_1\partial x_4} \\ \frac{\partial^2 f}{\partial x_2\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2\partial x_3} & \frac{\partial^2 f}{\partial x_2\partial x_4} \\ \frac{\partial^2 f}{\partial x_3\partial x_1} & \frac{\partial^2 f}{\partial x_3\partial x_2} & \frac{\partial^2 f}{\partial x_3^2} & \frac{\partial^2 f}{\partial x_3\partial x_4} \\ \frac{\partial^2 f}{\partial x_4\partial x_1} & \frac{\partial^2 f}{\partial x_4\partial x_2} & \frac{\partial^2 f}{\partial x_4\partial x_3} & \frac{\partial^2 f}{\partial x_4^2} \end{bmatrix}}$$

1. $\frac{\partial^2 f}{\partial x_1^2} = x_4 e^{-\frac{(x_1-1)^2}{2}}\left(\frac{(x_1-1)^2}{1} - 1\right)$
2. $\frac{\partial^2 f}{\partial x_2^2} = \frac{2(1-x_2^2)}{(x_2^2+1)^2} - \frac{1}{4(x_3^2+2)x_2^{3/2}}$
3. $\frac{\partial^2 f}{\partial x_3^2} = -\cos(x_3)\sin(x_4^2) + \frac{2\sqrt{x_2}(3x_3^2-2)}{(x_3^2+2)^3}$
4. $\frac{\partial^2 f}{\partial x_4^2} = -2x_4 e^{-\frac{(x_1-1)^2}{2}} + \cos(x_3)(-2\sin(x_4^2)\cdot 2x_4^2 + 2\cos(x_4^2))$
5. $\frac{\partial^2 f}{\partial x_1\partial x_2} = \frac{\partial^2 f}{\partial x_2\partial x_1} = 0$
6. $\frac{\partial^2 f}{\partial x_1\partial x_3} = \frac{\partial^2 f}{\partial x_3\partial x_1} = 0$:
7. $\frac{\partial^2 f}{\partial x_1\partial x_4} = \frac{\partial^2 f}{\partial x_4\partial x_1} = -(x_1-1)e^{-\frac{(x_1-1)^2}{2}}$
8. $\frac{\partial^2 f}{\partial x_2\partial x_3} = \frac{\partial^2 f}{\partial x_3\partial x_2} == -\frac{x_3}{(x_3^2+2)^2\sqrt{x_2}}$
9. $\frac{\partial^2 f}{\partial x_2\partial x_4} = \frac{\partial^2 f}{\partial x_4\partial x_2} = 0$
10. $\frac{\partial^2 f}{\partial x_3\partial x_4} = \frac{\partial^2 f}{\partial x_4\partial x_3} = -\cos(x_3)\cos(x_4^2)\cdot 2x_4$

So final Hessian Matrix is

$$Hf = \begin{pmatrix} x_4 e^{-\frac{(x_1-1)^2}{2}}((x_1-1)^2-1) & & & -(x_1-1)e^{-\frac{(x-1)^2}{2}} \\ & \frac{2(1-x_2^2)}{(x_2^2+1)^2} - \frac{1}{4(x_3^3+2)x_2^{3/2}} & -\frac{x_3}{(x_3^2+2)^2\sqrt{x_2}} & \\ & -\frac{x_3}{(x_3^2+2)^2\sqrt{x}} & -\cos(x_3)\sin(x_4^2) + \frac{2\sqrt{x_2}(3x_3^2+2)}{(x_3^2+2)^3} & -\cos(x_3)\cos(x_4^2)\,2x_4 \\ -(x_1-1)e^{-\frac{(x_1-1)^2}{2}} & & -\cos(x_3)\cos(x_4^2)\,2x_4 & -2x_4 e^{-\frac{(x_1-1)^2}{2}} + \cos(x_3)\cdot(-2\sin(x_4^2)\cdot 2x_4^2 + 2\cos(x_4^2)) \end{pmatrix}$$

Question 3

Let $\boldsymbol{A} \in \mathbb{R}^{p\times q}(p > q)$ be a matrix and has singular value decomposition

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$$

where $\boldsymbol{U} \in \mathbb{R}^{p\times q}, \boldsymbol{V} \in \mathbb{R}^{q\times q}$ are orthonormal matrices, and $\boldsymbol{D}$ is a diagonal matrix with nonnegative diagonal elements, i.e.,

$$\boldsymbol{D} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_q \end{bmatrix}$$

The diagonal elements are in a descending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q \geq 0$. Prove that

$$\sup_{\boldsymbol{x}\in\mathbb{R}^p, \|\boldsymbol{x}\|=1, \boldsymbol{y}\in\mathbb{R}^q, \|\boldsymbol{y}\|=1} \left|\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y}\right| = \lambda_1.$$

Answer

We can use the singular decomposition on $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top,$$

for any vectors $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{y} \in \mathbb{R}^q$, we can rewrite as $\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{y} = \boldsymbol{x}^\top\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{y}$. Define new vectors $\boldsymbol{u} = \boldsymbol{U}^\top\boldsymbol{x} \in \mathbb{R}^q$ and $\boldsymbol{v} = \boldsymbol{V}^\top\boldsymbol{y} \in \mathbb{R}^q$. Since $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal matrices, we have: $\|\boldsymbol{u}\| \leq \|\boldsymbol{x}\| = 1$ (because $\boldsymbol{U}$ is not necessarily a square), $\|\boldsymbol{v}\| = \|\boldsymbol{y}\| = 1$. Since we are proving the sup, we will choose the value of $\|\boldsymbol{u}\|$ to be its maximum, which is 1.

Thus, we can rewrite $\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{y} = \boldsymbol{u}^\top\boldsymbol{D}\boldsymbol{v} = \sum_{i=1}^q \lambda_i u_i v_i$. Since $\boldsymbol{D}$ is diagonal. By the Cauchy-Schwarz inequality, $|\sum_{i=1}^q \lambda_i u_i v_i| \leq \sum_{i=1}^q \lambda_i |u_i v_i|$.

Using the property that $\sum_{i=1}^q u_i^2 \leq 1$ and $\sum_{i=1}^q v_i^2 \leq 1$, the maximum value occurs when all the weight is concentrated on the largest singular value λ_1, this means that $u_1 = v_1 = 1$ and $u_i = v_i = 0$ for all $i \geq 2$. And thus we know that $\sup_{\|\boldsymbol{u}\|=1,\|\boldsymbol{v}\|=1} \sum_{i=1}^q \lambda_i u_i v_i = \lambda_1$

Then we choose $\boldsymbol{x} = \boldsymbol{u}_1$, $\boldsymbol{y} = \boldsymbol{v}_1$, where $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are the first columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively. Then we have $\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{y} = \boldsymbol{u}_1^\top\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{v}_1 = \lambda_1$.

Thus, the supremum is attained at λ_1, proving:

$$\sup_{\boldsymbol{x}\in\mathbb{R}^p, \|\boldsymbol{x}\|=1, \boldsymbol{y}\in\mathbb{R}^q, \|\boldsymbol{y}\|=1} \left|\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y}\right| = \lambda_1.$$

□