# BÀI TẬP VỀ NHÀ 2

MTH56-2425

Nguyễn Thị Mỹ Kim
MSSV: 23122040

Ngày 4 tháng 6 năm 2025

## 1

Gia đình phân phối chuẩn $\mathcal{D} : X \sim \mathcal{N}(\mu, \sigma^2)$ có hàm mật độ

$$\mathrm{p}(\mathrm{x}; \boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \boldsymbol{\alpha} = (\mu, \sigma) \in \mathbb{R} \times (0, \infty)$$

### a

Cho các mẫu dữ liệu độc lập và như nhau $\Theta = \{x_1, \ldots, x_n\}$. Hàm hợp lí của $p_{\boldsymbol{\alpha}}$ ứng với $\Theta$ là một hàm số theo $\boldsymbol{\alpha}$ được xác định bởi:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}|\Theta) &= \prod_{i=1}^{n} L(\boldsymbol{\alpha}|x_i) \\
&= \prod_{i=1}^{n} \mathrm{p}(x_i; \boldsymbol{\alpha}) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}^{n}} \exp\left[\sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2}\right]
\end{aligned}$$

### b

Như vậy, hàm log-hợp lí của $p_{\boldsymbol{\alpha}}$ ứng với $\Theta$:

$$
\begin{aligned}
l(\boldsymbol{\alpha}|\Theta) &= \ln\left(\prod_{i=1}^{n} L(\boldsymbol{\alpha}|x_i)\right) \\
&= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}^{n}} \exp\left[\sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2}\right]\right) \\
&= \frac{n}{2}\ln\left(\frac{1}{2\pi\sigma^2}\right) + \sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2} \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2}
\end{aligned}
$$

**c**

Tập xác định của $\boldsymbol{\alpha} = (\mu, \sigma)$ là $\mathbb{R} \times (0, \infty)$. Mà $\mathbb{R}$ là tập lồi (mọi đoạn thẳng nối hai điểm thuộc $\mathbb{R}$ đều nằm trong $\mathbb{R}$) và $(0, \infty)$ là tập lồi (mọi đường thẳng nối hai số dương đều nằm trong $(0, \infty)$). Do đó tích Descartes của $\mathbb{R}$ và $(0, \infty)$) là tập lồi. Ta lại có:

- $p_{\boldsymbol{\alpha}}$: họ phân phối tham số hóa bởi $\boldsymbol{\alpha} \in \mathcal{D}$
- $\Theta = (x_1, \dots, x_n)$: các giá trị được giả sử lấy mẫu từ $p_{\boldsymbol{\alpha}}$

Tham số ước lượng theo phương pháp MLE:

$$\boldsymbol{\alpha}_{MLE} = \text{argmax}_{\boldsymbol{\alpha} \in \mathcal{D}} \mathcal{L}(\boldsymbol{\alpha}|\Theta)$$

hay ta sử dụng hàm đồng biến log

$$\boldsymbol{\alpha}_{MLE} = \text{argmax}_{\boldsymbol{\alpha} \in \mathcal{D}} \log \mathcal{L}(\boldsymbol{\alpha}|\Theta) = \text{argmax}_{\boldsymbol{\alpha} \in \mathcal{D}} l(\boldsymbol{\alpha}|\Theta)$$

Hàm $l(\mu, \sigma|\Theta)$ là tổ hợp của các hàm sơ cấp như $\ln(\cdot)$, $(\cdot)^2$, và $1/(\cdot)$, đều là các hàm khả vi bậc hai trên miền $\mathbb{R} \times (0, \infty)$. Do đó, $l \in C^2(\mathbb{R} \times (0, \infty))$. Khi đó, hàm cũng khả vi bậc một trên miền này.
Đạo hàm hàm $l(\boldsymbol{\alpha}|\Theta)$ từng phần theo $\mu$ và $\sigma$:
Khi đó ta có:

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n} \frac{(x_i - \mu)}{\sigma^2} = \frac{1}{\sigma^2}(-n\mu + \sum_{i=1}^{n} x_i)$$

và

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2$$

Ta tìm điểm dừng bằng cách giải hệ:

$$\begin{cases} \frac{\partial l}{\partial \mu} = 0 \\ \frac{\partial l}{\partial \sigma} = 0 \end{cases}$$
$$\Leftrightarrow \begin{cases} \mu = \frac{\sum_{i=1}^n x_i}{n} \\ \sigma^2 = \frac{\sum_{i=1}^n (x_i-\mu)^2}{n} \end{cases}$$

Ta tìm ma trận Hessian của $l(\boldsymbol{\alpha}|\Theta)$:

$$H = \nabla^2 l(\mu, \sigma|\Theta) = \begin{bmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu\,\partial \sigma} \\ \frac{\partial^2 l}{\partial \sigma\,\partial \mu} & \frac{\partial^2 l}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{2}{\sigma^3}\sum_{i=1}^n (x_i-\mu) \\ -\frac{2}{\sigma^3}\sum_{i=1}^n (x_i-\mu) & \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^n (x_i-\mu)^2 \end{bmatrix}$$

Xét $\nabla^2 l(\mu, \sigma|\Theta)$ tại điểm dừng $\boldsymbol{\alpha}^* = (\mu, \sigma) = (\frac{\sum_{i=1}^n x_i}{n}, \sqrt{\frac{\sum_{i=1}^n (x_i-\mu)^2}{n}})$, khi đó:

- $-\frac{1}{\sigma^4}\sum_{i=1}^n (x_i-\mu) = 0$
- $\frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^n (x_i-\mu)^2 = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} n\sigma^2 = -\frac{2n}{\sigma^2}$

Như vậy, ma trận Hessian lúc này trở thành:

$$H = \begin{bmatrix} -\frac{n}{\sigma^2} & 0 \\ 0 & -\frac{2n}{\sigma^2} \end{bmatrix} \preceq 0$$

H lúc này là một ma trận xác định âm.
Như vậy, với $l(\boldsymbol{\alpha}|\Theta)$ là hàm khả vi bậc hai và $\nabla l(\boldsymbol{\alpha}^*|\Theta) = 0$ và $\nabla^2 l(\boldsymbol{\alpha}^*|\Theta) \preceq 0$ thì $\boldsymbol{\alpha}^*$ là điểm cực đại địa phương của hàm $l(\boldsymbol{\alpha}|\Theta)$. Do $\boldsymbol{\alpha}^*$ là điểm dừng duy nhất của hàm $l(\boldsymbol{\alpha}|\Theta)$, nên hàm cũng không tồn tại cực đại khác. Do đó, $\boldsymbol{\alpha}^*$ là cực đại toàn cục.
Vậy $\boldsymbol{\alpha}^* = (\mu, \sigma) = (\frac{\sum_{i=1}^n x_i}{n}, \sqrt{\frac{\sum_{i=1}^n (x_i-\mu)^2}{n}})$ là nghiệm của bài toán tối ưu.

# 2

## a

Xét tập dữ liệu $\Theta = \{x_1, \ldots, x_n\} \subset \mathbb{R}^2$ và tham số h. Hàm kernel $K(u)$ là phân phối chuẩn để ước lượng mật độ kernel $\hat{f}(x)$ tại điểm $x$:

$$\hat{f}(x, \Theta) = \frac{1}{nh^2}\sum_{j=1}^n K(\frac{x - x_j}{h}), \text{ với } K(u) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}\|u\|^2\right)$$

Công thức cập nhật:

$$x_i^{(t+1)} = \frac{\sum_{j=1}^{n=4} x_j K(x_i^{(t)} - x_j)}{\sum_{j=1}^{n=4} K(x_i^{(t)} - x_j)}$$

$\Theta$ là tập các điểm dữ liệu ban đầu: $\Theta = \{x_j\}$ với $j \in \{1,2,3,4\}$. Ta khởi tạo tập các điểm ban đầu để bắt đầu thuật toán Mean Shift là $\Sigma = \{x_i^{(0)}\}$, trong đó mỗi điểm $x_i^{(0)}$ được gán bằng chính điểm $x_j$ tương ứng trong $\Theta$, tức là $x_i^{(0)} = x_j$ với $i = j$.

| t | i | $\sum_{j=1}^{n=4} x_j K(x_i^{(t)} - x_j)$ | $\sum_{j=1}^{n=4} K(x_i^{(t)} - x_j)$ | $x_i^{(t+1)}$ |
|---|---|---|---|---|
| t=0 | 1 | $\begin{pmatrix}0.08410174\\0.96332007\end{pmatrix}$ | 2.5515663725129367 | $\begin{pmatrix}0.03296083\\0.37754067\end{pmatrix}$ |
| | 2 | $\begin{pmatrix}0.08410174\\1.5882463\end{pmatrix}$ | | $\begin{pmatrix}0.03296083\\0.62245933\end{pmatrix}$ |
| | 3 | $\begin{pmatrix}7.57059738\\0.96332007\end{pmatrix}$ | | $\begin{pmatrix}2.96703917\\0.37754067\end{pmatrix}$ |
| | 4 | $\begin{pmatrix}7.57059738\\1.5882463\end{pmatrix}$ | | $\begin{pmatrix}2.96703917\\0.62245933\end{pmatrix}$ |
| t=1 | 1 | $\begin{pmatrix}0.10137352\\1.30931058\end{pmatrix}$ | 2.789189182693252 | $\begin{pmatrix}0.03634516\\0.46942337\end{pmatrix}$ |
| | 2 | $\begin{pmatrix}0.10137352\\1.4798786\end{pmatrix}$ | | $\begin{pmatrix}0.03634516\\0.53057663\end{pmatrix}$ |
| | 3 | $\begin{pmatrix}8.26619402\\1.30931058\end{pmatrix}$ | | $\begin{pmatrix}2.96365484\\0.46942337\end{pmatrix}$ |
| | 4 | $\begin{pmatrix}8.26619402\\1.4798786\end{pmatrix}$ | | $\begin{pmatrix}2.96365484\\0.53057663\end{pmatrix}$ |

Bảng 1: Hai bước đầu tiên của thuật toán Mean Shift với Gaussian kernel

## b

Chứng minh rằng mỗi bước cập nhật mean shift tương ứng với một bước của thuật toán gradient ascent cho hàm ước lượng mật độ kernel tại thời điểm t. Viết ra hàm tốc độ học tương ứng.

Ta xét hàm $\hat{f}$ với một giá trị trong $\Sigma$ là tham số $x_i^{(t)}$ ($x_i$ tại thời điểm $t$)

$$\hat{f}(x_i^{(t)}, \Theta) = \frac{1}{nh^2} \sum_{j=1}^{n} \frac{1}{2\pi} \exp\left(-\frac{1}{2} \| \frac{x_i^{(t)} - x_j}{h} \|^2\right) \text{ với } K(\frac{x_i^{(t)} - x_j}{h}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} \| \frac{x_i^{(t)} - x_j}{h} \|^2\right)$$

Khi đó:

$$\begin{aligned}
\nabla \hat{f}(x_i^{(t)}, \Theta) &= \frac{1}{nh^2} \sum_{j=1}^{n} \frac{-1}{2} \frac{2(x_i^{(t)} - x_j)}{h^2} \frac{1}{2\pi} \exp\left( -\frac{1}{2} \| \frac{x_i^{(t)} - x_j}{h} \|^2 \right) \\
&= \frac{-1}{nh^4} \sum_{j=1}^{n} (x_i^{(t)} - x_j) K(\frac{x_i^{(t)} - x_j}{h}) \\
&= \frac{1}{nh^4} \sum_{j=1}^{n} x_j K(\frac{x_i^{(t)} - x_j}{h}) - x_i^{(t)} \frac{1}{nh^4} \sum_{j=1}^{n} K(\frac{x_i^{(t)} - x_j}{h})
\end{aligned}$$

Ta có công thức gradient ascent là $x_i^{(t+1)} = x_i^{(t)} + \eta \nabla \hat{f}(x_i^{(t)}, \Theta)$ với $\eta$ là learning rate (tốc độ học). Như vậy:

$$\begin{aligned}
x_i^{(t+1)} &= x_i^{(t)} + \eta \left( \frac{1}{nh^4} \sum_{j=1}^{n} x_j K(\frac{x_i^{(t)} - x_j}{h}) - x_i^{(t)} \frac{1}{nh^4} \sum_{j=1}^{n} K(\frac{x_i^{(t)} - x_j}{h}) \right) \\
&= x_i^{(t)} - \eta \frac{1}{nh^4} x_i^{(t)} \sum_{j=1}^{n} K(\frac{x_i^{(t)} - x_j}{h}) + \eta \frac{1}{nh^4} \sum_{j=1}^{n} x_j K(\frac{x_i^{(t)} - x_j}{h})
\end{aligned}$$

Nhận xét: $x_i^{(t)} - \eta \frac{1}{nh^4} x_i^{(t)} \sum_{j=1}^{n} K(\frac{x_i^{(t)} - x_j}{h})$ là một giá trị không đổi tại lần cập nhật thứ $(t)$ cho $x_i$. Vì thế ta có thể chọn:

$$\eta = \frac{nh^4}{\sum_{j=1}^{n} K(\frac{x_i^{(t)} - x_j}{h})}$$

để $x_i^{(t)} - \eta \frac{1}{nh^4} x_i^{(t)} \sum_{j=1}^{n} K(\frac{x_i^{(t)} - x_j}{h}) = 0$.
Như vậy:

$$x_i^{(t+1)} = \frac{\sum_{j=1}^{n} x_j K(\frac{x_i^{(t)} - x_j}{h})}{\sum_{j=1}^{n} K(\frac{x_i^{(t)} - x_j}{h})} \text{ với } \eta = \frac{nh^4}{\sum_{j=1}^{n} K(\frac{x_i^{(t)} - x_j}{h})}$$