

Bài tập về nhà 2

Toán học cho Trí tuệ nhân tạo

16 Tháng 5, 2025

Lưu ý: Việc **chép đáp án** từ nguồn có sẵn như bài của bạn khác hoặc trên mạng internet mà **không hiểu** sẽ nhận điểm 0.

Ước lượng hợp lý cực đại (MLE) cho phân phối chuẩn

Hàm phân phối xác suất của phân phối chuẩn được tham số hóa bởi trung bình μ và độ lệch chuẩn σ như sau:

$$p(x; \boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right], \quad \boldsymbol{\alpha} = (\mu, \sigma).$$

- (a) (2 điểm) Cho mẫu dữ liệu độc lập và như nhau $\Theta = \{x_1, \dots, x_n\}$. Hãy viết ra hàm hợp lý $L(\boldsymbol{\alpha} | \Theta)$ của một phân phối chuẩn ứng với mẫu dữ liệu trên.
- (b) (1 điểm) Hãy viết ra hàm log-hợp lý $\ell(\boldsymbol{\alpha} | \Theta)$.
- (c) (2 điểm) Chứng minh rằng ước lượng hợp lý cực đại $\boldsymbol{\alpha}_{MLE}$ là trung bình của mẫu dữ liệu:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

và căn bậc hai của phương sai của mẫu dữ liệu:

$$Var = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Thuật toán phân cụm mean shift và gradient ascent

Thuật toán **mean shift** là một kỹ thuật phân cụm dựa trên ước lượng mật độ kernel (KDE). Cho tập dữ liệu $\Theta = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^2$ và tham số h cho trước. Ta sử dụng hàm kernel $K(u)$ là

phân phối chuẩn để xây dựng hàm ước lượng mật độ kernel $\hat{f}(x)$ tại điểm x như sau:

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

Ở bước thứ t , $t = 1, 2, \dots$, thuật toán cập nhật $\Theta^{(t)}$ với mọi i theo công thức mean shift:

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

Thuật toán dừng khi các điểm $x_i^{(t)}$ hội tụ.

Yêu cầu:

(a) (2 điểm) Thực hiện 2 bước của thuật toán mean shift trên tập dữ liệu sau với $h = 1$:

$$\Theta = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}.$$

(b) (3 điểm) 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.

Ghi chú:

- Việc lựa chọn hàm kernel $K(u)$ là phân phối chuẩn là để thuận tiện cho việc tính đạo hàm. Tồn tại những lựa chọn khác.
- So với thuật toán gradient descent cổ điển, phiên bản gradient descent của mean shift có tốc độ học lẫn hàm tối ưu hóa thay đổi theo từng bước t .