

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP
THÀNH PHỐ HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN

ĐỀ THI PRETEST HK1 2024-2025

Lớp/Lớp học phần: ĐHKHMT16-17

Ngày thi: 20/12/2024

Thời gian làm bài: 60 phút

Họ và tên thí sinh; MSSV:; STT.....

Phần 1: Bắt buộc

Câu 1: (3,0 điểm)

Cho một mẫu ngẫu nhiên kích thước $n=3$, $(X_1; X_2; X_3)$ được chọn từ tổng thể có trung bình μ và phương sai σ^2 , với giá trị $a \in \mathbb{R}$ đặt:

$$T_a = \frac{1}{4}((X_1 + aX_2 + X_3))$$

1. Tìm a sao cho T_a là ước lượng không chệch của μ .

(Gợi ý: $\hat{\theta}$ gọi là ước lượng không chệch của μ nếu $\mathbb{E}(\hat{\theta}) = \mu$)

2. Tính phương sai T_a .

(Gợi ý: $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \mathbb{E}(X_3) = \mu$ và $\text{Var}(X_1) = \text{Var}(X_2) = \text{Var}(X_3) = \sigma^2$)

3. Xét trung bình bình phương sai số $\text{MSE}(T_a, \mu) = \mathbb{E}[(T_a - \mu)^2]$. Tìm giá trị a sao cho T_a là ước lượng tốt nhất theo nghĩa cực tiểu hoá trung bình bình phương sai số.

Câu 2: (2,0 điểm)

Cho class A bao gồm 09 mẫu có giá trị: $A = \{1, 2, 3, 3, 4, 4, 6, 6, 8\}$.

Cho class B bao gồm 09 mẫu có giá trị: $B = \{4, 6, 7, 7, 8, 9, 9, 10, 12\}$.

Tìm biên phân lớp của 2 lớp trên sao cho tổng lỗi bé nhất

Phần 2: Sinh viên chọn 1 trong 2 câu

Câu 1: (5,0 điểm)

Giả sử dữ liệu được thu thập từ 2 lớp với các giá trị đặc trưng 1 chiều x như sau:

- Lớp ω_1 : $D_1 = \{2, 2, 3, 3, 4, 4, 5, 5, 11, 11, 11, 12, 12, 13, 13, 14, 15, 16\}$;
- Lớp ω_2 : $D_2 = \{9, 9, 10, 10, 11, 14, 14, 14, 14, 15, 15, 16, 16, 18, 19\}$.

- a) Viết công thức hàm cửa sổ Parzen $p_\phi(x) = ?$

- b) Sử dụng hàm cửa sổ Parzen $\phi(u) = \begin{cases} 1, & \text{nếu } |u| \leq \frac{1}{2} \\ 0, & \text{ngược lại} \end{cases}$ để xác định $p_\phi(11|\omega_1)$ và

$p_\phi(11|\omega_2)$ với kích thước cửa sổ $h = 5$.

- c) Phân lớp đối tượng với giá trị đặc trưng $x = 11$.

Câu 2: (5,0 điểm)

1. Giả sử một mạng nơ ron có 3 lớp với 17 nút ở lớp nhập, 25 nút ở lớp ẩn, 10 nút ở lớp xuất. Cho biết kích thước ma trận trọng số:
 - Lớp nhập và lớp ẩn?
 - Lớp ẩn và lớp xuất?
2. Trình bày một số phương pháp giảm overfitting của mạng nơ ron mà bạn biết?.
3. Trong các bài toán dưới đây bạn sử dụng thuật toán logistic regression cho bài toán nào là phù hợp.
 - Phân loại spam email
 - Dự đoán giá nhà dựa trên các thuộc tính căn nhà (diện tích, số phòng)
 - Phân loại bệnh nhân có bệnh hoặc không bị bệnh
4. Giả sử áp dụng thuật toán SVM trong phân lớp ta thu được siêu phẳng có phương trình $w^T x + b = 0$, với $w = [1, -2]$, $b = 3$. Cho 3 điểm $x_1 = [2, 1]$, $x_2 = [-0.5, 1.5]$, $x_3 = [-1.75, 0.5]$. Tính toán và xác định nhãn đầu ra (+1 hoặc -1) của giá trị tương ứng $y_1 = ?$, $y_2 = ?$, $y_3 = ?$
5. Thông thường khi huấn luyện mạng nơ ron cho bài toán phân lớp (số lớp > 2). Hàm loss được sử dụng là hàm nào?. Giải thích?

Ghi chú:

- Sinh viên không được phép sử dụng tài liệu, cán bộ coi thi không giải thích gì thêm.
- Sinh viên lưu bài theo định dạng: STT_MaSoSinhVien_HoTenSinhVienKhongDau

----- Hết -----