



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ 1 – Năm học 2024-2025

MÃ LƯU TRỮ
(do phòng KT-ĐBCL ghi)
CK2425.1
MTH10322

Tên học phần: Nhận dạng mẫu Mã HP: MTH10322
Thời gian làm bài: 90 phút Ngày thi: 14/01/2025
Ghi chú: *Sinh viên không được phép sử dụng tài liệu khi làm bài.*



Câu 1 (3 điểm).

Xét tập dữ liệu sau:

Weather	Temperature	Play?
Sunny	Hot	No
Sunny	Hot	No
Overcast	Hot	Yes
Rainy	Mild	Yes
Rainy	Cool	Yes
Rainy	Cool	No

Chúng ta quan tâm đến bài toán phân loại nhị phân, với các thuộc tính (features) là 2 cột Weather và Temperature, và nhãn (label) là cột Play?.

- (1 điểm) Tính *entropy* của nhãn Play?.
- (0.75 điểm) Tính *Information Gain (IG)* nếu sử dụng thuộc tính Weather để chia tập dữ liệu.
- (0.75 điểm) Tính *Information Gain (IG)* nếu sử dụng thuộc tính Temperature để chia tập dữ liệu.
- (0.5 điểm) Giải thích cách chọn thuộc tính tốt nhất dựa trên *IG* trong thuật toán cây quyết định.

Câu 2 (2 điểm).

- (1 điểm) Chứng minh rằng:

$$H(X, Y) = H(X) + H(Y|X),$$

- (1 điểm) Chứng minh rằng:

$$H(X|Y) \leq H(X),$$

với dấu "=" xảy ra khi X và Y độc lập.

(Đề thi gồm 2 trang)

Họ tên người ra đề/MSCB:..... Chữ ký: [Trang 1/2]
Họ tên người duyệt đề:..... Chữ ký:



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ 1 – Năm học 2024-2025

MÃ LƯU TRỮ
(do phòng KT-ĐBCL ghi)
CK2425.1
MTH10322

Câu 3 (2 điểm).

(2 điểm) Cho mô hình hồi quy tuyến tính đa biến:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \epsilon,$$

trong đó: $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{w} \in \mathbb{R}^p$, và $\epsilon \in \mathbb{R}^n$ (giả định $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$).
Với hàm mất mát dựa trên sai số bình phương:

$$L(\mathbf{w}) = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 = (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}).$$

Hãy tìm hệ số hồi quy $\mathbf{w}^*$ bằng cách tối thiểu hóa hàm mất mát $L(\mathbf{w})$.

Câu 4 (2.5 điểm).

Xét bài toán tối ưu sau:

$$\max_{\mathbf{u}} \mathbf{u}^\top \mathbf{C} \mathbf{u}, \quad \text{với ràng buộc } \|\mathbf{u}\| = 1,$$

trong đó:

- $\mathbf{C}$ là ma trận đối xứng kích thước $n \times n$,
- $\mathbf{u} \in \mathbb{R}^n$.

1. (1 điểm) Chứng minh nghiệm của bài toán này chính là vector riêng (eigenvector) tương ứng với trị riêng (eigenvalue) lớn nhất của ma trận $\mathbf{C}$.
2. (1.5 điểm) Giải thích bài toán này liên quan như thế nào đến phương pháp phân tích thành phần chính (Principal Component Analysis - PCA).

Câu 5 (2.5 điểm).

1. (1 điểm) Trình bày thuật toán Random Forest.
2. (0.5 điểm) Giải thích sự khác nhau giữa Random Forest và Bagging.
3. (1 điểm) Trình bày thuật toán Gradient Boosting.

(Đề thi gồm 2 trang)

Họ tên người ra đề/MSCB: Chữ ký:
Họ tên người duyệt đề: Chữ ký:

[Trang 2/2]