

# BÀI KIỂM TRA ĐẦU GIỜ

## Xác suất có điều kiện & Định lý Bayes

*Thời gian: 15 phút / Không được sử dụng tài liệu*

Họ tên: \_\_\_\_\_

Mã SV: \_\_\_\_\_

### PHẦN LÝ THUYẾT (4 điểm)

#### Câu 1 (2 điểm): Điền vào chỗ trống

- a) Xác suất có điều kiện  $P(A|B)$  được định nghĩa là:  $P(A|B) =$  \_\_\_\_\_  
b) Định lý Bayes có công thức:  $P(A|B) =$  \_\_\_\_\_  
c) Phân phối Bernoulli có tham số  $p$ , kỳ vọng  $E[X] =$  \_\_\_\_\_ và phương sai  $\text{Var}(X) =$  \_\_\_\_\_

#### Câu 2 (2 điểm): Câu hỏi lý thuyết

a) Tại sao định lý Bayes lại quan trọng trong học máy? Cho một ví dụ cụ thể về việc sử dụng định lý Bayes để cập nhật xác suất.

b) Giải thích sự khác biệt giữa phân phối Bernoulli và phân phối Chuẩn. Cho ví dụ thực tế cho mỗi phân phối.

### PHẦN BÀI TẬP (6 điểm)

#### Câu 3 (3 điểm): Bài toán phân loại email spam

Một hệ thống lọc email có thông tin sau:

- 30% email là spam
- Từ "khuyến mãi" xuất hiện trong 80% email spam
- Từ "khuyến mãi" xuất hiện trong 10% email bình thường

a) Tính xác suất một email chứa từ "khuyến mãi".

b) Nếu một email chứa từ "khuyến mãi", tính xác suất email đó là spam.

c) Nếu bạn thiết kế một classifier đơn giản dựa trên từ "khuyến mãi" (phân loại là spam nếu chứa từ này), accuracy của classifier này là bao nhiêu?

#### **Câu 4 (3 điểm): Ứng dụng Bernoulli trong A/B Testing**

Một website muốn test hiệu quả của nút "Mua ngay" mới. Với nút cũ, tỷ lệ click là 0.15. Với nút mới, sau 100 lần hiển thị, có 20 lần click.

a) Mô hình hóa số lần click bằng phân phối nào? Tại sao?

b) Tính xác suất có đúng 20 lần click trong 100 lần hiển thị nếu tỷ lệ click thực sự là 0.15.

c) Dựa trên kết quả trên, bạn có kết luận gì về hiệu quả của nút mới?