

# Basic Python - Bayes Theory

*Hoàng-Nguyên Vũ*

## 1 Lý Thuyết về Naive Bayes

Naive Bayes là một thuật toán phân loại dựa trên lý thuyết xác suất. Thuật toán này giả định rằng các đặc trưng đầu vào là độc lập với nhau khi lớp đầu ra đã được biết. Thuật toán được gọi là "naive" vì giả định này có thể không đúng trong thực tế, nhưng thường vẫn hiệu quả.

### 1.1 Định Lý Bayes

Định lý Bayes mô tả xác suất của một sự kiện dựa trên thông tin về các sự kiện khác. Công thức của định lý Bayes là:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Trong đó:

- $P(C|X)$  là xác suất của lớp  $C$  khi biết các đặc trưng  $X$ .
- $P(X|C)$  là xác suất của các đặc trưng  $X$  khi lớp là  $C$ .
- $P(C)$  là xác suất tiên nghiệm của lớp  $C$ .
- $P(X)$  là xác suất của các đặc trưng  $X$ .

### 1.2 Giả Định Độc Lập

Naive Bayes giả định rằng các đặc trưng là độc lập với nhau trong mỗi lớp. Điều này có nghĩa là:

$$P(X|C) = P(x_1, x_2, \dots, x_n|C) = \prod_{i=1}^n P(x_i|C) \quad (2)$$

Trong đó,  $x_i$  là các đặc trưng riêng lẻ.

### 1.3 Ưu Điểm và Nhược Điểm

**Ưu Điểm:**

- Đơn giản và dễ triển khai.
- Hiệu quả với nhiều đặc trưng.
- Thường hoạt động tốt với dữ liệu có kích thước lớn và thiếu dữ liệu.

**Nhược Điểm:**

- Giả định các đặc trưng độc lập có thể không chính xác trong nhiều tình huống thực tế.
- Có thể kém hiệu quả nếu các đặc trưng thực tế không độc lập.

## 2 Bài Tập Toán về Naive Bayes

**Dữ Liệu:** Một cửa hàng có hai loại sản phẩm: Sản phẩm A và Sản phẩm B. Trong số 100 sản phẩm, 60 là sản phẩm A và 40 là sản phẩm B. Các đặc trưng của sản phẩm bao gồm màu sắc (Đỏ hoặc Xanh) và kích thước (Nhỏ hoặc Lớn). Dữ liệu như sau:

- Trong số 60 sản phẩm A, 20 sản phẩm là Đỏ và Nhỏ, 10 sản phẩm là Đỏ và Lớn, 15 sản phẩm là Xanh và Nhỏ, 15 sản phẩm là Xanh và Lớn.
- Trong số 40 sản phẩm B, 10 sản phẩm là Đỏ và Nhỏ, 5 sản phẩm là Đỏ và Lớn, 10 sản phẩm là Xanh và Nhỏ, 15 sản phẩm là Xanh và Lớn.

### 2.1 Yêu Cầu

1. Tính xác suất tiên nghiệm của mỗi lớp (Sản phẩm A và Sản phẩm B).
2. Tính xác suất điều kiện của màu sắc và kích thước trong mỗi lớp.
3. Dự đoán lớp của một sản phẩm mới có đặc trưng: Đỏ và Nhỏ.

### 2.2 Đáp án

#### 1. Xác Suất Tiên Nghiệm

$$P(A) = 0.6 \quad (3)$$

$$P(B) = 0.4 \quad (4)$$

#### 2. Xác Suất Điều Kiện

- Cho lớp A:

$$P(\text{Đỏ và Nhỏ}|A) \approx 0.333 \quad (5)$$

$$P(\text{Đỏ và Lớn}|A) \approx 0.167 \quad (6)$$

$$P(\text{Xanh và Nhỏ}|A) = 0.25 \quad (7)$$

$$P(\text{Xanh và Lớn}|A) = 0.25 \quad (8)$$

- Cho lớp B:

$$P(\text{Đỏ và Nhỏ}|B) = 0.25 \quad (9)$$

$$P(\text{Đỏ và Lớn}|B) = 0.125 \quad (10)$$

$$P(\text{Xanh và Nhỏ}|B) = 0.25 \quad (11)$$

$$P(\text{Xanh và Lớn}|B) = 0.375 \quad (12)$$

### 3. Dự Đoán Lớp

Sử dụng định lý Bayes để tính xác suất của lớp A và lớp B cho một sản phẩm mới có đặc trưng Đỏ và Nhỏ:

$$P(A|\text{Đỏ và Nhỏ}) = 0.1998 \quad (13)$$

$$P(B|\text{Đỏ và Nhỏ}) = 0.1 \quad (14)$$

**Kết luận:** Với xác suất cao hơn, sản phẩm mới có khả năng thuộc lớp A.