

# Probability Exercise

(Naive Bayes Classifier)

Ngày 2 tháng 10 năm 2022

## 1. Binay Classification - PLAY TENNIS

Cho tập dữ liệu huấn luyện mô hình phân loại nhị phân Naive Bayes gồm các thuộc tính "Out look", "Temperature", "Humidity", "Wind":

| Day | Outlook  | Temperature | Humidity | Wind   | PlayTennis |
|-----|----------|-------------|----------|--------|------------|
| D1  | Sunny    | Hot         | High     | Weak   | No         |
| D2  | Sunny    | Hot         | High     | Strong | No         |
| D3  | Overcast | Hot         | High     | Weak   | Yes        |
| D4  | Rain     | Mild        | High     | Weak   | Yes        |
| D5  | Rain     | Cool        | Normal   | Weak   | Yes        |
| D6  | Rain     | Cool        | Normal   | Strong | No         |
| D7  | Overcast | Cool        | Normal   | Strong | Yes        |
| D8  | Overcast | Mild        | High     | Weak   | No         |
| D9  | Sunny    | Cool        | Normal   | Weak   | Yes        |
| D10 | Rain     | Mild        | Normal   | Weak   | Yes        |

Bảng 1: Play Tennis - Tập dữ liệu huấn luyện

**Cho sự kiện thử nghiệm:**

**X = (Outlook=Sunny, Temperature=Cool, Humidity=High, Wind=Strong)**

1.1 Xác suất xảy ra sự kiện "Play Tennis"="Yes" và sự kiện "Play Tennis"="No" lần lượt là:

- (A)  $P(\text{"Play Tennis"} = \text{"Yes"}) = 6/10$ ,  $P(\text{"Play Tennis"} = \text{"No"}) = 4/10$
- (B)  $P(\text{"Play Tennis"} = \text{"Yes"}) = 4/10$ ,  $P(\text{"Play Tennis"} = \text{"No"}) = 6/10$
- (C)  $P(\text{"Play Tennis"} = \text{"Yes"}) = 6/10$ ,  $P(\text{"Play Tennis"} = \text{"No"}) = 6/10$
- (D)  $P(\text{"Play Tennis"} = \text{"Yes"}) = 4/10$ ,  $P(\text{"Play Tennis"} = \text{"No"}) = 4/10$

1.2 Xác suất xảy ra sự kiện "Play Tennis"="Yes" khi sự kiện X xảy ra là:

- (A)  $P(\text{"Play Tennis"} = \text{"Yes"} | X) \propto 0.0014$
- (B)  $P(\text{"Play Tennis"} = \text{"Yes"} | X) \propto 0.0028$
- (C)  $P(\text{"Play Tennis"} = \text{"Yes"} | X) \propto 0.0188$
- (D)  $P(\text{"Play Tennis"} = \text{"Yes"} | X) \propto 0.0098$

1.3 Xác suất xảy ra sự kiện "Play Tennis"="No" khi sự kiện X xảy ra là:

- (A)  $P(\text{"Play Tennis"} = \text{"No"} | X) \propto 0.0014$
- (B)  $P(\text{"Play Tennis"} = \text{"No"} | X) \propto 0.0028$

$$(C) P(\text{"Play Tennis"} = \text{"No"} \mid X) \propto 0.0188$$

$$(D) P(\text{"Play Tennis"} = \text{"No"} \mid X) \propto 0.0098$$

1.4 Khi xảy ra sự kiện X, nhân của "Play Tennis" sẽ là:

$$(A) \text{"Play Tennis"} = \text{"Yes"}$$

$$(B) \text{"Play Tennis"} = \text{"No"}$$

## 2. Multi-label Classification - TRAFFIC DATA

Cho tập dữ liệu huấn luyện mô hình phân loại Naive Bayes gồm các thuộc tính "Day", "Season", "Fog", "Rain":

| Day      | Season | Fog    | Rain   | Class     |
|----------|--------|--------|--------|-----------|
| Weekday  | Spring | None   | None   | On Time   |
| Weekday  | Winter | None   | Slight | On Time   |
| Weekday  | Winter | None   | None   | On Time   |
| Holiday  | Winter | High   | Slight | Late      |
| Saturday | Summer | Normal | None   | On Time   |
| Weekday  | Autumn | Normal | None   | Very Late |
| Holiday  | Summer | High   | Slight | On Time   |
| Sunday   | Summer | Normal | None   | On Time   |
| Weekday  | Winter | High   | Heavy  | Very Late |
| Weekday  | Summer | None   | Slight | On Time   |
| Saturday | Spring | High   | Heavy  | Cancelled |
| Weekday  | Summer | High   | Slight | On Time   |
| Weekday  | Winter | Normal | None   | Late      |
| Weekday  | Summer | High   | None   | On Time   |
| Weekday  | Winter | Normal | Heavy  | Vary Late |
| Saturday | Autumn | High   | Slight | On Time   |
| Weekday  | Autumn | None   | Heavy  | On Time   |
| Holiday  | Spring | Normal | Slight | On Time   |
| Weekday  | Spring | Normal | None   | On Time   |
| Weekday  | Spring | Normal | Heavy  | On Time   |

Bảng 2: Traffic Data - Tập dữ liệu huấn luyện

Cho sự kiện thử nghiệm:

$X = (\text{Day}=\text{Weekday}, \text{Season}=\text{Winter}, \text{Fog}=\text{High}, \text{Rain}=\text{Heavy})$

2.1 Xác suất xảy ra sự kiện "Class"="One Time", sự kiện "Class"="Late", sự kiện "Class"="Very Late" và sự kiện "Class"="Cancelled" lần lượt là:

$$(A) P(\text{"Class"} = \text{"On Time"}) = 14/20, P(\text{"Class"} = \text{"Late"}) = 2/20,$$

$$P(\text{"Class"} = \text{"Very Late"}) = 3/20, P(\text{"Class"} = \text{"Cancelled"}) = 1/20$$

$$(B) P(\text{"Class"} = \text{"On Time"}) = 2/20, P(\text{"Class"} = \text{"Late"}) = 3/20,$$

$$P(\text{"Class"} = \text{"Very Late"}) = 1/20, P(\text{"Class"} = \text{"Cancelled"}) = 14/20$$

$$(C) P(\text{"Class"} = \text{"On Time"}) = 3/20, P(\text{"Class"} = \text{"Late"}) = 1/20,$$

$$P(\text{"Class"} = \text{"Very Late"}) = 2/20, P(\text{"Class"} = \text{"Cancelled"}) = 14/20$$

$$(D) P(\text{"Class"} = \text{"On Time"}) = 1/20, P(\text{"Class"} = \text{"Late"}) = 1/20,$$

$$P(\text{"Class"} = \text{"Very Late"}) = 14/20, P(\text{"Class"} = \text{"Cancelled"}) = 3/20$$

2.2 Xác suất xảy ra sự kiện “Class”=”One Time” khi sự kiện X xảy ra là:

- (A)  $P(\text{“Class”} = \text{“On Time”} \mid X) \propto 0.0222$
- (B)  $P(\text{“Class”} = \text{“On Time”} \mid X) \propto 0.0013$
- (C)  $P(\text{“Class”} = \text{“On Time”} \mid X) \propto 0.0026$
- (D)  $P(\text{“Class”} = \text{“On Time”} \mid X) \propto 0.0000$

2.3 Xác suất xảy ra sự kiện “Class”=”Late” khi sự kiện X xảy ra là:

- (A)  $P(\text{“Class”} = \text{“Late”} \mid X) \propto 0.0222$
- (B)  $P(\text{“Class”} = \text{“Late”} \mid X) \propto 0.0013$
- (C)  $P(\text{“Class”} = \text{“Late”} \mid X) \propto 0.0026$
- (D)  $P(\text{“Class”} = \text{“Late”} \mid X) \propto 0.0000$

2.4 Xác suất xảy ra sự kiện “Class”= ”Very Late” khi sự kiện X xảy ra là:

- (A)  $P(\text{“Class”} = \text{“Very Late”} \mid X) \propto 0.0222$
- (B)  $P(\text{“Class”} = \text{“Very Late”} \mid X) \propto 0.0013$
- (C)  $P(\text{“Class”} = \text{“Very Late”} \mid X) \propto 0.0026$
- (D)  $P(\text{“Class”} = \text{“Very Late”} \mid X) \propto 0.0000$

2.5 Xác suất xảy ra sự kiện “Class”= Cancelled” khi sự kiện X xảy ra là:

- (A)  $P(\text{“Class”} = \text{“Cancelled”} \mid X) \propto 0.0222$
- (B)  $P(\text{“Class”} = \text{“Cancelled”} \mid X) \propto 0.0013$
- (C)  $P(\text{“Class”} = \text{“Cancelled”} \mid X) \propto 0.0026$
- (D)  $P(\text{“Class”} = \text{“Cancelled”} \mid X) \propto 0.0000$

2.6 Khi sự kiện X xảy ra, vậy “Class” sẽ có nhãn là:

- (A) “Class” = “On Time”
- (B) “Class” = “Late”
- (C) “Class” = “Very Late”
- (D) “Class” = “Cancelled”

**3. Cho một tập dữ liệu huấn luyện phân loại hoa Iris dựa vào chiều dài cánh hoa như bảng dữ liệu bên dưới. Các bạn hãy trả lời các câu hỏi sau khi dùng Gaussian Naive Bayes cho data Iris này.**

|        |     |     |     |     |     |     |     |     |     |     |     |     |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Length | 1.4 | 1.0 | 1.3 | 1.9 | 2.0 | 1.8 | 3.0 | 3.8 | 4.1 | 3.9 | 4.2 | 3.4 |
| Class  | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   | 1   | 1   | 1   | 1   |

Bảng 3: Phân loại cánh hoa Iris dựa vào chiều dài cánh hoa - Tập dữ liệu huấn luyện

3.1 Giá trị mean và variance của biến đầu vào (Length) cho “Class”=“0” lần lượt là:

- (A) mean = 1.566 và variance = 0.128
- (B) mean = 3.733 và variance = 0.172
- (C) mean = 1.566 và variance = 0.172

3.2 Giá trị mean và variance của biến đầu vào (Length) cho “Class”=“1” lần lượt là:

- (A) mean = 1.566 và variance = 0.128
- (B) mean = 3.733 và variance = 0.172
- (C) mean = 1.566 và variance = 0.172

3.3 Cho dữ liệu kiểm thử  $X = (\text{Length}=3.4)$ . Xác suất dữ liệu kiểm thử  $X$  thuộc vào “Class”=“0” và “Class”=“1” lần lượt là:

- (A)  $P(\text{“Class”} = \text{“0”} \mid X) = 0.000003$  và  $P(\text{“Class”} = \text{“1”} \mid X) = 0.999997$
- (B)  $P(\text{“Class”} = \text{“0”} \mid X) = 0.999992$  và  $P(\text{“Class”} = \text{“1”} \mid X) = 0.000008$
- (C)  $P(\text{“Class”} = \text{“0”} \mid X) = 0.000008$  và  $P(\text{“Class”} = \text{“1”} \mid X) = 0.999992$