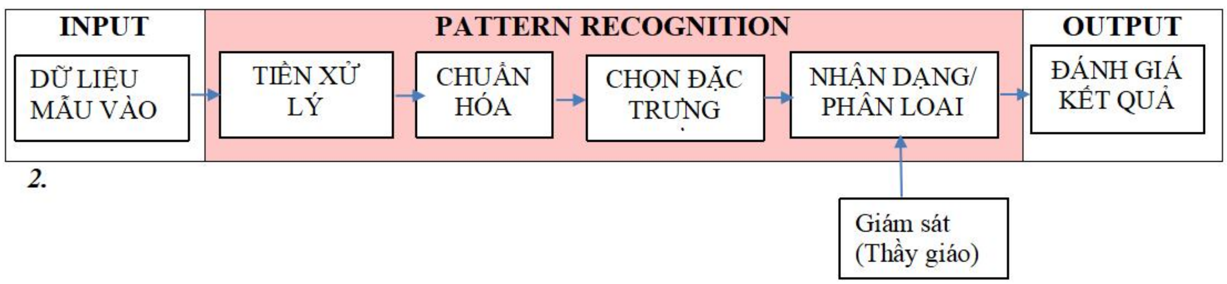
CÂU HỎI MÔN PATTERN RECOGNITION

ĐỀ 1

Câu 1

a) **Vẽ sơ đồ cấu trúc hệ thống nhận dạng mẫu.**



b) **Nêu rõ ý nghĩa, đặc điểm và ứng dụng cho các đối tương nhận mẫu khác nhau?**

- Tiếng nói: Đặc điểm gồm tần số cơ bản và các formant giúp phân biệt các âm thanh và giọng nói. Ứng dụng trong nhận dạng giọng nói và các hệ thống trợ lý ảo.

- Hình ảnh: Các đặc trưng như độ sáng, hình dạng, hoặc đường biên ảnh. Ứng dụng trong phân loại hình ảnh, nhận diện khuôn mặt.

- Văn bản: Sử dụng các đặc trưng như từ vựng và ngữ pháp. Ứng dụng trong phân loại văn bản, phát hiện thư rác, dịch máy.

- Mạng nơ-ron sâu (CNN): Dùng trong trích xuất đặc trưng của hình ảnh qua các lớp tích chập để phát hiện các chi tiết hình ảnh. Phù hợp với các bài toán phức tạp như nhận dạng vật thể và xử lý ảnh y tế

Câu 2. Nêu ba phương pháp phân biệt để giải quyết các bài toán quyết định (Thống kê)

Phân loại bằng lý thuyết quyết định Bayes

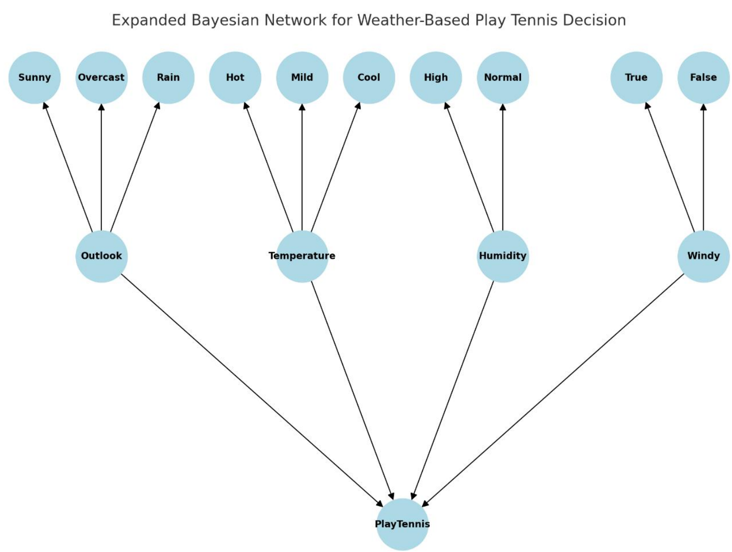
Hàm phân biệt

Phân phối chuẩn

ĐỀ 2

Câu 1. Cho bài toán thời tiết

a) Vẽ mạng Bayes



b) Viết thuật toán theo bước tính xác suất của Temperature trong điều kiện Play Tennis?

Cho t = {Hot, Mild, Cool}; k = {Yes, No}

+ Duyệt lần lượt giá trị k:

    + Duyệt lần lượt giá trị t:

P(Temperature=t | PlayTennis=Yes)=P(Temperature=t, PlayTennis=Yes)/P(PlayTennis=Yes)

Trong đó:

P(PlayTennis = k): Tính bằng tổng số lần PlayTennis là k chia cho tổng số mẫu.

P(Temperature = t, PlayTennis = Yes): Số lần xuất hiện giá trị cụ thể t của Temperature khi PlayTennis = k chia cho tổng số mẫu mà PlayTennis = k

Câu 2. Nêu các phương pháp nhận mẫu thống kê chính

Phân loại bằng lý thuyết quyết định Bayes

Phân biệt tuyến tính

Phân biệt phi tuyến

Phân loại dựa trên xác suất và Naive Bayes

Phân loại k-láng giềng gần nhấtCây quyết định

ĐỀ 3

Câu 1. Thế nào là suy diễn tiến trong mạng Bayes. Hãy cho ví dụ từ bài toán thời tiết

Suy diễn Bayesian là một kiểu suy diễn đơn giản mà máy tính dựa vào các xác suất của dữ liệu đã có trước của bài toán để tìm ra xác suất của đầu ra cho những đầu vào tiếp theo.

Giả sử ta biết các thông tin:

- **Outlook = Sunny**
- **Temperature = Hot**
- **Humidity = High**
- **Windy = False**

Tính xác suất **PlayTennis = Yes** trong các điều kiện này.

P(PlayTennis=Yes | Outlook=Sunny,Temperature=Hot,Humidity=High,Windy=False)

= P(PlayTennis=Yes | Outlook=Sunny)*P(PlayTennis=Yes | Temperature=Hot)

*P(PlayTennis=Yes | Humidity=High)*P(PlayTennis=Yes | Windy=False)

Câu 2. Nhận mẫu dựa trên thống kê học gồm mấy giai đoạn:

Mô tả và viết công thức toán mô tả các giai đoạn đó

1. **Tiền xử lý (Preprocessing)**: Chuẩn hóa và biến đổi dữ liệu để giảm nhiễu và chuẩn bị dữ liệu cho các bước tiếp theo.
   - Ví dụ: Chuẩn hóa biến X thành $X' = \frac{X - \text{mean}}{\text{std}}$.
2. **Trích xuất đặc trưng (Feature Extraction)**: Chọn ra các đặc trưng quan trọng để giảm chiều dữ liệu.
   - Ví dụ: Lấy trung bình hoặc phương sai của dữ liệu: $\text{Feature} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
3. **Phân loại (Classification)**: Dùng mô hình xác suất để phân loại mẫu vào các lớp.
   - Ví dụ: Sử dụng lý thuyết Bayes $P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$.

ĐỀ 4

Câu 1. Thế nào là suy diễn lùi trong mạng Bayes. Hãy cho ví dụ từ bài toán thời tiết.

Giả sử ta đã quan sát thấy rằng **PlayTennis = Yes** và muốn tính xác suất của một nguyên nhân tiềm ẩn, ví dụ: **Outlook = Sunny**.

Khi đó:

P(Outlook=Sunny | PlayTennis=Yes)

[P(PlayTennis=Yes | Outlook=Sunny)×P(Outlook=Sunny)]

= --

P(PlayTennis=Yes)

Câu 2: Chứng minh bằng ví dụ qua bài toán luật tổng của xác suất có điều kiện

Giả sử ta muốn tính xác suất có thể chơi tennis, không phân biệt điều kiện thời tiết nào. Các điều kiện về **Outlook** có thể là "Sunny", "Overcast", và "Rain".

Theo luật tổng, xác suất **Play Tennis = Yes** được tính bằng:

$$P(\text{Play Tennis} = \text{Yes}) = P(\text{Yes}|\text{Sunny}) \cdot P(\text{Sunny}) + P(\text{Yes}|\text{Overcast}) \cdot P(\text{Overcast}) + P(\text{Yes}|\text{Rain}) \cdot P(\text{Rai}$$

Bằng cách tính từng phần tử dựa trên dữ liệu, ta có thể xác định xác suất tổng của việc có thể chơi tennis bất kể điều kiện thời tiết.

ĐỀ 5

Câu 1.

a) Chứng minh định lý Bayes

Giả sử ta có hai sự kiện A và B, với xác suất chung của chúng được xác định bằng quy tắc nhân:

$$P(A \wedge B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

Từ đó, ta có thể viết:

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Định lý Bayes cho phép tính xác suất của A khi biết B, dựa trên xác suất của B khi biết A và xác suất ban đầu của A và B.

c) Khi nào thì áp dụng định lý bayes, khi nào dùng định nghĩa để tính xác suất có điều kiện. Cho ví dụ minh họa

- **Áp dụng định lý Bayes** khi bạn có một biến điều kiện và cần suy luận ngược lại để tìm xác suất của biến hỏi (nguyên nhân), thường trong trường hợp có các xác suất biên và xác suất điều kiện của các biến khác nhau

Dùng định nghĩa xác suất có điều kiện khi chỉ cần xác suất xảy ra của một sự kiện trong điều kiện một sự kiện khác đã xảy ra, mà không có yêu cầu suy luận ngược.

Ví dụ:

1. **Định lý Bayes**: Xác suất một email là **thư rác (Spam)** dựa trên sự xuất hiện của từ "khuyến mãi" trong email.

$$P(\text{Spam}|\text{Khuyến mãi}) = \frac{P(\text{Khuyến mãi}|\text{Spam}) \cdot P(\text{Spam})}{P(\text{Khuyến mãi})}$$

2. **Định nghĩa xác suất có điều kiện**: Tính xác suất trời **nắng** trong điều kiện đã biết là **chơi tennis**:

$$P(\text{Sunny}|\text{Play Tennis} = \text{Yes}) = \frac{P(\text{Sunny} \wedge \text{Play Tennis} = \text{Yes})}{P(\text{Play Tennis} = \text{Yes})}$$

Câu 2. Thế nào là xác suất hậu nghiệm: Viết công thức, nêu tên và ý nghĩa của các thành phần của công thức

Xác suất hậu nghiệm là xác suất của một sự kiện sau khi đã biết một số thông tin hoặc bằng chứng mới. Đây là xác suất được tính toán sau khi đã cập nhật các thông tin liên quan, thường dùng trong định lý Bayes

Công thức xác suất hậu nghiệm:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Ý nghĩa các thành phần:

- $P(A|B)$: Xác suất hậu nghiệm của A khi biết B (sau khi cập nhật bằng chứng B).
- $P(B|A)$: Xác suất có điều kiện của B khi biết A (còn gọi là xác suất khả năng).
- $P(A)$: Xác suất tiên nghiệm của A (trước khi biết bằng chứng B).
- $P(B)$: Xác suất biên của B (tổng xác suất của bằng chứng).

ĐỀ 6

Câu 1

a) Thế nào là hàm phân biệt (Discriminant function)

Hàm phân biệt là một hàm giúp xác định và gán một vector đầu vào x cho một trong các lớp Ck . Trong trường hợp đơn giản, hàm phân biệt là một hàm tuyến tính với đầu vào, thường được sử dụng trong các bài toán phân loại, với bề mặt quyết định là các siêu mặt phẳng trong không gian đầu vào

b) Viết hàm phân biệt và công thức tính (Ví dụ cho bài toán thời tiết)

$$y(x) = w^T x + w_0$$

Câu 2: Trinh bày suy diễn từ dưới lên của mạng Bayes dưới dạng thuật toán

Thuật toán suy diễn từ dưới lên trong mạng Bayes

- Khởi tạo xác suất cho các nút lá:
- Truyền thông tin từ các nút lá lên các nút cha
- Tính xác suất hậu nghiệm cho các nút cha
- Lặp lại cho đến khi đạt tới gốc của mạng

- Kết luận và cập nhật xác suất hậu nghiệm

ĐỀ 7

Câu 1. Cho $p(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)p(\mathcal{C}_k)}{p(\mathbf{x})}$ **.**

a) Cho biết tên gọi của các thành phần

Ck: biến hỏi
X: biến điều kiện

- P(Ck|x): Xác suất của sự kiện Ck xảy ra khi biết rằng sự kiện x đã xảy ra (xác suất hậu nghiệm).
- P(x|Ck): Xác suất của sự kiện x xảy ra khi biết rằng sự kiện Ck đã xảy ra (xác suất điều kiện).
- P(Ck): Xác suất của biến hỏi Ck xảy ra (xác suất tiên nghiệm).
- P(x): Xác suất của điều kiện x xảy ra (xác suất tổng quát).

b) Nêu cách tính p(x) dùng phương pháp chuẩn hóa trong mạng Bayes

Giả sử chúng ta có một mạng Bayes đơn giản với các biến A và B, và chúng ta biết các xác suất có điều kiện $P(A|B)$ và $P(B)$. Chúng ta muốn tính $P(A)$.

1. **Xác định các xác suất có điều kiện:**
   - $P(A|B)$
   - $P(B)$
2. **Tính xác suất kết hợp:**
   - $P(A,B) = P(A|B) \cdot P(B)$
3. **Tính tổng xác suất:**
   - $P(A) = \sum_B P(A,B) = \sum_B P(A|B) \cdot P(B)$
4. **Chuẩn hóa xác suất:**
   - $P(A) = \frac{P(A,B)}{\sum_B P(A,B)} = \frac{P(A|B) \cdot P(B)}{\sum_B P(A|B) \cdot P(B)}$

Câu 2. Hãy cho biết trong bài toán thời tiết, chúng ta sử dụng phương pháp thứ mấy trong ba phương pháp đưa ra quyết định

- Phương pháp thống kê : Sử dụng lý thuyết quyết định Bayes

ĐỀ 8

Câu 1

a) Thế nào là học trong mô hình thống kê

- Trong mô hình thống kê, "học" chính là quá trình ước lượng tham số để mô hình phù hợp với phân phối dữ liệu nhất. Điều này còn gọi là **ước lượng tham số** (parameter estimation), nhằm xác định các tham số của mô hình thống kê dựa trên dữ liệu có sẵn.

b) Viết công thức tìm tham số tổng quát. Cho Ví dụ Tìm Tham số B: Play Tennis)

$$\theta = \arg\max_{\theta} \prod_{i=1}^{N} p(x_i|\theta)$$

Câu 2. Nêu các tính chất của xác suất có điều kiện (hay các luật)

1) **Qui tắc hay luật nhân (Product Rule)**

$$P(A, B)=P(A \land B)=P(A|B)*P(B)$$

2) **Qui tắc chuỗi (là mở rộng của quy tắc nhân)**

$$P(A, B, C, D)=P(A \land B \land C \land D)=P(A|B, C, D)*P(B|C, D)*P(C|D)*P(D)$$

3) **Quy tắc chuỗi có điều kiện**

$$P(A, B|C)=P(A|B, C)*P(B|C)$$

4) **Quy tắc cộng (Sum Rule)**

$$P(A) = \sum_b P(A|B = b) * P(B = b)$$

ĐỀ 9

Câu 1

a) Thế nào là xác suất biên (Marginal Probability). Cho ví dụ xác suất biên từ bài toán Thời tiết?

Xác suất biên là xác suất xảy ra của một biến cụ thể mà không quan tâm đến các biến khác trong không gian xác suất. Nó được tính bằng cách tổng hoặc tích phân xác suất có điều kiện của các biến còn lại.

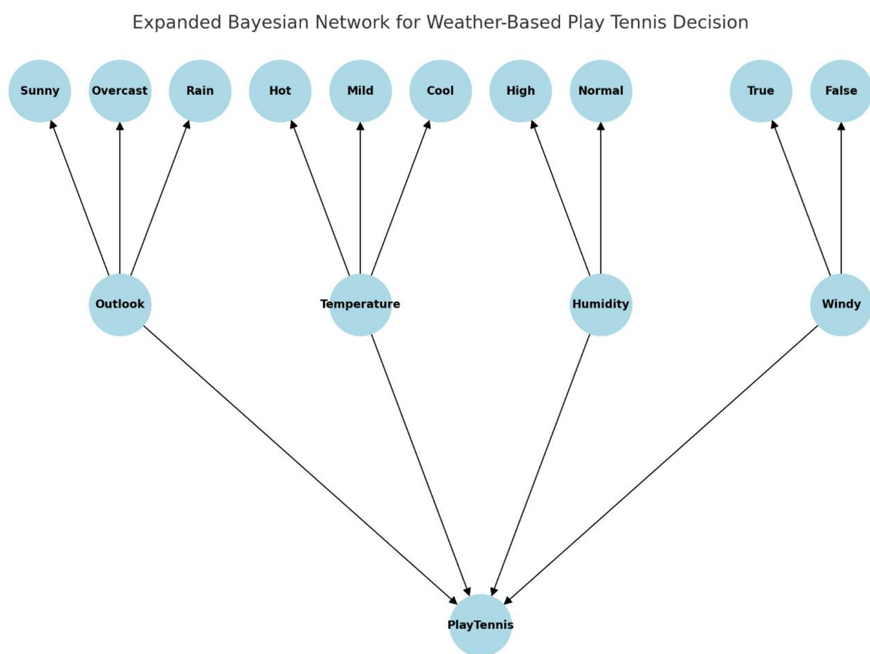
-Ví dụ từ bài toán Thời tiết: Nếu xác suất để thời tiết "Nắng" là P(Sunny) đây là một xác suất biên vì nó không phụ thuộc vào các yếu tố khác như độ ẩm hoặc gió.

b) Xác suất biên chính là luật gì trong các luật (hay tính chất) của xác suất có điều kiện?

Xác suất biên là kết quả của **quy tắc cộng** trong các tính chất xác suất có điều kiện. Theo quy tắc này, xác suất của một biến ngẫu nhiên được tính bằng tổng xác suất có điều kiện của các trường hợp có thể xảy ra của các biến khác trong không gian xác suất.

Câu 2: Định nghĩa mạng Bayes. Cho ví dụ

Mạng Bayes (Bayesian Network) là một mô hình đồ thị xác suất biểu diễn tập hợp các biến ngẫu nhiên và các mối quan hệ có điều kiện giữa chúng. Mạng Bayes sử dụng lý thuyết Bayes để tính toán xác suất của các biến dựa trên các quan sát và các mối quan hệ có điều kiện.



ĐỀ 10

Câu 1.

a) **Cho một chuỗi 8 ngày liên tục với 3 trạng thái Nắng, mưa, râm tùy chọn. Lập công thức tính xác suất để có 8 ngày như vây. Loại xác suất đó gọi là xác suất gì?**

Chuỗi markov chain
=> P(O| Model) = P(1,2,3,4,5,6,7,8|Model)
= P(2|1)*P(3|2)*(4|3)*P(5|4)*P(6|5)*P(7|6)*P(8|7)*P(8)
= a12 * a23 * a34 * a45 * a56 * a67 * a78 *P(8)

b) Thế nào là ma trận chuyển trạng thái? Điều kiện của các trạng thái đó? Cho ví dụ bảng trạng thái với ba trạng thái Năng, Mưa, Râm

Ví dụ: Với các trạng thái Nắng, Mưa, và Râm, ma trận chuyển trạng thái có thể như sau:

$$\begin{pmatrix} P(\text{Nắng}|\text{Nắng}) & P(\text{Mưa}|\text{Nắng}) & P(\text{Râm}|\text{Nắng}) \\ P(\text{Nắng}|\text{Mưa}) & P(\text{Mưa}|\text{Mưa}) & P(\text{Râm}|\text{Mưa}) \\ P(\text{Nắng}|\text{Râm}) & P(\text{Mưa}|\text{Râm}) & P(\text{Râm}|\text{Râm}) \end{pmatrix}$$

Ví dụ cụ thể:

$$\begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.5 & 0.2 & 0.3 \end{pmatrix}$$

Ma trận này cho biết xác suất chuyển đổi giữa các trạng thái thời tiết theo từng ngày.

- Đặc điểm: xác suất trên cùng 1 hàng bằng 1

Câu 2. Viết công thức tính xác suất Bayes tổng quát. Nêu rõ ý nghĩa các tham số trong công thức đó thông qua một ví dụ (Ví dụ mạng Bayes)

$$p(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)p(\mathcal{C}_k)}{p(\mathbf{x})}$$

Ck: biến hỏi
X: biến điều kiện

- P(Ck|x): Xác suất của sự kiện Ck xảy ra khi biết rằng sự kiện x đã xảy ra (xác suất hậu nghiệm).
- P(x|Ck): Xác suất của sự kiện x xảy ra khi biết rằng sự kiện Ck đã xảy ra (xác suất điều kiện).
- P(Ck): Xác suất của biến hỏi Ck xảy ra (xác suất tiên nghiệm).
- P(x): Xác suất của điều kiện x xảy ra (xác suất tổng quát).