TOÁN ỨNG DỤNG CNTT

CHUYÊN ĐỀ XÁC SUẤT

SV: TRƯƠNG QUANG LỘC

PHẦN 1

BIẾN NGẪU NHIÊN

01 - BIẾN NGẪU NHIÊN RỜI RẠC

02 - BIẾN NGẪU NHIÊN LIÊN TỤC

PHẦN 1

BIẾN NGẪU NHIÊN

KHÁI NIỆM

Biến ngẫu nhiên: là 1 đại lượng có thể nhận giá trị này hay giá trị khác tùy thuộc vào kết quả của phép thử ngẫu nhiên

- BNN rời rạc: là biến ngẫu nhiên có hữu hạn hoặc vô hạn đếm được các giá trị
- BNN liên tục: là biến ngẫu nhiên có các giá trị lấp đầy một khoảng số thực

TÍNH CHẤT

- Tổng, hiệu, tích, thương của các BNN là một BNN
- BNN hằng: hằng số được coi là BNN

PHẦN 1

BIẾN NGẪU NHIÊN

Hàm phân phối xác suất của BNN rời rạc

tính tổng xác suất của tất cả các giá trị nhỏ hơn hoặc bằng một giá trị cụ thể trong miền giá trị của biến cố

TÍNH CHẤT

- $0 \le F(x) \le 1$
- $x1 < x2$ thì $F(x1) \le F(x2)$
- $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$
- $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$
- $P(a \le X \le b) = F(b) - F(a)$

VD: Tung ngẫu nhiên 3 đồng xu. Gọi X là số mặt sấp xuất hiện.

$F(1) = P(X<1) = P(X=0) = P(\{NNN\}) = 1/8$

$F(0.3) = P(X<0.3) = P(X=0) = P(\{NNN\}) = 1/8$

$F(-1) = P(X<-1) = 0$

PHẦN 1 **BIẾN NGẪU NHIÊN**

Hàm mật độ xác suất của BNN liên tục

Hàm mật độ xác suất được biểu diễn bằng một đồ thị liên tục (một đường biểu diễn) trong thống kê xác định tỷ lệ phân phối xác suất (khả năng xảy ra một kết quả) cho một biến ngẫu nhiên rời rạc.

Cho X là BNNLT có hàm phân phối xác suất F(X). Nếu tồn tại hàm số f(x) không âm, khả tích trên R, f(x) là hàm mật độ xác suất của BNNLT X khi:

$$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad x \in \mathbb{R}}$$

PHẦN 1 BIẾN NGẪU NHIÊN

Hàm mật độ xác suất của BNN liên tục

TÍNH CHẤT

1. $f_X(x) \geq 0$ với mọi $x \in \mathbb{R}$.

2. $P(a < X < b) = \int_a^b f_X(x)dx$.

3. $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

PHẦN 1

BIẾN NGẪU NHIÊN

Hàm mật độ xác suất của BNN liên tục

TÍNH CHẤT

1. $f_X(x) \geq 0$ với mọi $x \in \mathbb{R}$.

2. $P(a < X < b) = \int_a^b f_X(x)dx$.

3. $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

PHẦN 2 XÁC SUẤT ĐỒNG THỜI

Hàm phân phối xác suất đồng thời hay hàm phân phối tích luỹ xác suất đồng thời (*Joint CDF - Joint Cumulative Probability Distribution Function*) của 2 biến ngẫu nhiên X, Y được định nghĩa như sau:

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) \quad , x, y \in \mathbb{R}$$

XÁC SUẤT ĐỒNG THỜI BNNRR

$$p_{X,Y}(x,y) = P(X = x, Y = y)$$

TÍNH CHẤT

- $0 \leq p(x_i, y_j) \leq 1$
- $\sum\limits_{\forall i} \sum\limits_{\forall j} p(x_i, y_j) = 1$
- $F(x,y) = \sum\limits_{\forall x_i \leq x} \sum\limits_{\forall y_j \leq y} p(x_i, y_j)$

PHẦN 2

XÁC SUẤT ĐỒNG THỜI

XÁC SUẤT ĐỒNG THỜI BNNLT

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y)$$

hay

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)\,du\,dv$$

TÍNH CHẤT

- $f(x,y) > 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$
- $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x,y)\,dx\,dy$
- $P(X = x, Y = y) = \int_{y}^{y} \int_{x}^{x} f(x,y)\,dx\,dy = 0$

PHẦN 3

XÁC SUẤT BIÊN

Phân phối biên/Xác suất biên (Marginal Probability) là phân phối của riêng từng biến một.

$$\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(X \leq x, Y < +\infty) \\
&= F_{X,Y}(x, +\infty) \\
F_Y(y) &= P(Y \leq y) \\
&= P(+\infty, Y \leq y) \\
&= F_{X,Y}(+\infty, y)
\end{aligned}$$

PHẦN 3

XÁC SUẤT BIÊN

XÁC SUẤT BIÊN BNNRR

$$
\begin{aligned}
p_X(x) &= P(X = x) \\
&= \sum_{\forall j} p(x, y_j) \\
p_Y(y) &= P(Y = y) \\
&= \sum_{\forall i} p(x_i, y)
\end{aligned}
$$

XÁC SUẤT BIÊN BNNRR

$$
\begin{aligned}
f_X(x) &= P(X = x) \\
&= \int_{-\infty}^{\infty} f(x, y)dy \\
&= \frac{\partial}{\partial x} F_X(x) \\
f_Y(y) &= P(Y = y) \\
&= \int_{-\infty}^{\infty} f(x, y)dx \\
&= \frac{\partial}{\partial y} F_Y(y)
\end{aligned}
$$

PHẦN 4

XÁC SUẤT CÓ ĐIỀU KIỆN

Xác suất của biến cố Y với điều kiện biến cố X đã xảy ra

X, Y RỜI RẠC

$$\begin{aligned} p_{X|Y}(x|y) &= P(X=x|Y=y) \\ &= \frac{P(X=x, Y=y)}{P(Y=y)} \\ &= \frac{p_{X,Y}(x,y)}{p_Y(y)} \\ &= \frac{p_{X,Y}(x,y)}{\sum_{\forall i} p(x_i, y)} \end{aligned}$$

X, Y LIÊN TỤC

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy} \end{aligned}$$

PHẦN 5

BIẾN NGẪU NHIÊN ĐỘC LẬP

2 biến X, Y độc lập khi xác suất của chúng không phụ thuộc vào nhau. Như ta đã biết 2 sự kiện A, B độc lập khi và chỉ khi P(AB) = P(A)P(B), tương tự với BNN chúng độc lập khi và chỉ khi:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad ,\forall x,y \in \mathbb{R}$$

X, Y RỜI RẠC

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad ,\forall x,y \in \mathbb{R}$$

X, Y LIÊN TỤC

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad ,\forall x,y \in \mathbb{R}$$

PHẦN 6

CÁC ĐẠI LƯỢNG ĐẶC TRƯNG

KỲ VỌNG

X, Y RỜI RẠC

$$
\begin{aligned}
E[X] &= \sum_{i} x_i p(x_i) \\
&= \sum_{\forall i} x_i \sum_{\forall j} p(x_i, y_j) \\
&= \sum_{\forall i} \sum_{\forall j} x_i p(x_i, y_j)
\end{aligned}
$$

X, Y LIÊN TỤC

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) dx dy
\end{aligned}
$$

PHẦN 6

CÁC ĐẠI LƯỢNG ĐẶC TRƯNG

PHƯƠNG SAI

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= E[X^2] - E^2[X] \end{aligned}$$

HIỆP PHƯƠNG SAI

Hiệp phương sai (*Covariance*) của 2 biến ngẫu nhiên X, Y kí hiệu là $Cov(X, Y)$ được định nghĩa rằng:

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Tương tự như cách khai triển của phương sai, ta cũng sẽ thu được công thức tương đương sau:

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

PHẦN 6 CÁC ĐẠI LƯỢNG ĐẶC TRƯNG

HIỆP PHƯƠNG SAI

Hiệp phương sai có một số tính chất sau:

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, X) = Var(X)$
- $Cov(aX, bY) = abCov(X, Y)$ với a, b là hằng số
- $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$
- $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i}\sum_{j>i} Cov(X_i, X_j)$

Hiệp phương sai thường được tập hợp lại thành 1 ma trận đối xứng gọi là ma trận hiệp phương sai:

$$\begin{bmatrix} Cov(X,X) & Cov(X,Y) \\ Cov(Y,X) & Cov(Y,Y) \end{bmatrix} = \begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$

PHẦN 7

QUY TẮC BAYES

CÔNG THỨC

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

PHẦN 7

QUY TẮC BAYES

VÍ DỤ

Dây chuyền lắp ráp máy vô tuyến điện gồm các linh kiện là sản phẩm từ 2 nhà máy sản xuất ra. Số linh kiện nhà máy 1 sản xuất chiếm 55%, số linh kiện nhà máy 2 sản xuất chiếm 45%; tỷ lệ sản phẩm đạt tiêu chuẩn của nhà máy 1 là 90%, nhà máy 2 là 87%. Lấy ngẫu nhiên ra 1 linh kiện từ dây chuyền lắp ráp đó ra kiểm tra thì được kết quả linh kiện đạt chuẩn. Tìm xác suất để linh kiện đó do nhà máy 1 sản xuất?

Gọi A_i = "linh kiện do nhà máy thứ i sản xuất", $i = 1, 2$.

Gọi B = "linh kiện đạt chuẩn", ta cần tìm $P(A_1|B)$.

Ta có:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) = 0,55.0,9 + 0,45.0,87 = 0,8865.$$

Và

$$P(A_1/B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{0,55.0,9}{0,8865} = 0,5583.$$

PHẦN 8

PHÂN BỐ RỜI RẠC

PHÂN PHỐI BERNOULLI

Một phép thử Bernoulli có kết quả nhận được là một trong hai giá trị hoặc "thành công" hoặc "thất bại". "Thành công" xảy ra với xác suất là p, "thất bại" với xác suất là q=1-p.

Hàm xác suất: f1) = p, f(0) = 1 - p

Kỳ vọng: E(X) = p

Phương sai: σ^2 = p(1-p) = pq

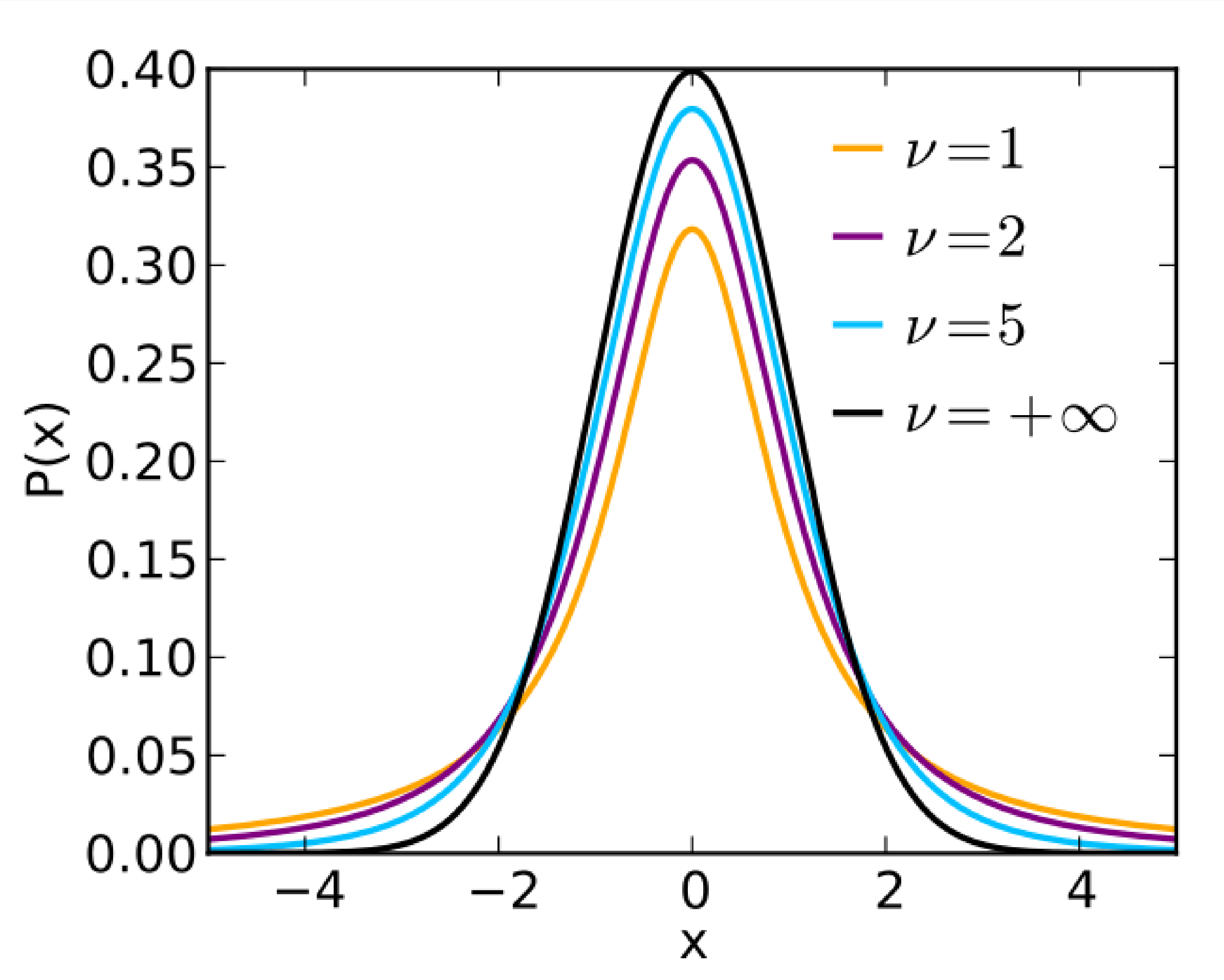
PHÂN PHỐI CHUẨN (GAUSSIAN)

Hàm mật độ xác suất của phân phối chuẩn với trung bình μ và phương sai σ^2 là một ví dụ của hàm Gauss

Ký hiệu: X ~ N(μ, σ^2)

E(X) = μ, V(X) = σ^2

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$



PHẦN 10 CHUỖI MARKOV

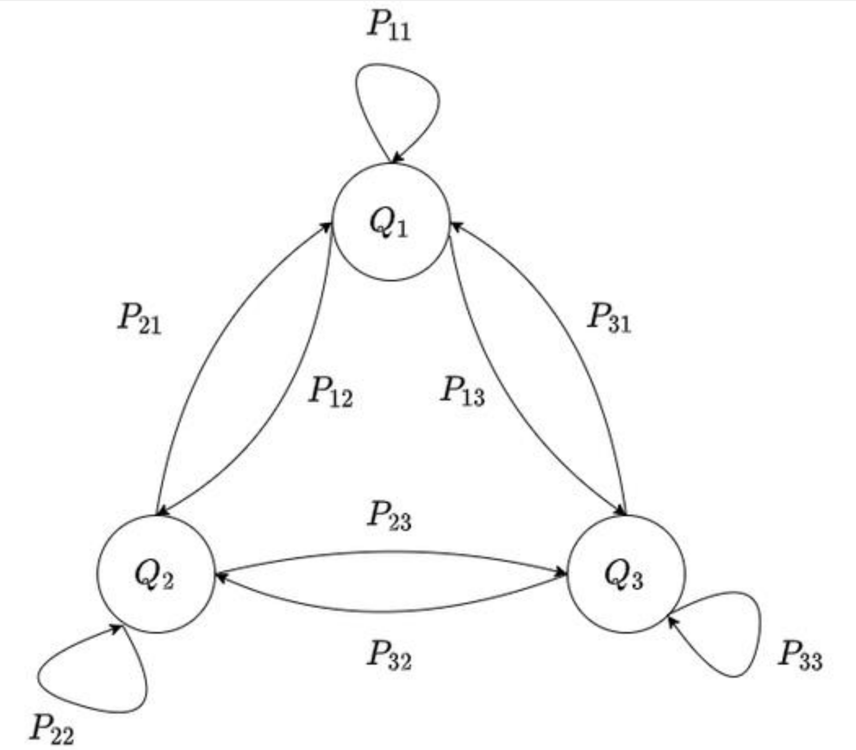
KHÁI NIỆM

là một mô hình ngẫu nhiên hay tiến trình ngẫu nhiên mô tả một chuỗi các sự kiện có khả năng xảy ra, mà xác suất để xảy ra sự kiện tiếp theo phụ thuộc chỉ vào sự kiện hiện tại.

là một mô hình "không có trí nhớ" (memorylessness), nghĩa là các sự kiện xảy ra trong quá khứ sẽ không được ghi nhớ, sự kiện trong tương lai chỉ phụ thuộc sự kiện hiện tại của mô hình.

là sự kết hợp bởi 2 thành phần: tập trạng thái Q và ma trận chuyển đổi giữa các trạng thái P

có hình dạng trông giống như một đồ thị có hướng, có trọng số, với các node trên đồ thị là các trạng thái trong tập trạng thái của Markov chain, và các trọng số trên các cạnh của đồ thị mô tả xác suất chuyển Pij, là xác suất di chuyển từ trạng thái Qi đến Qj



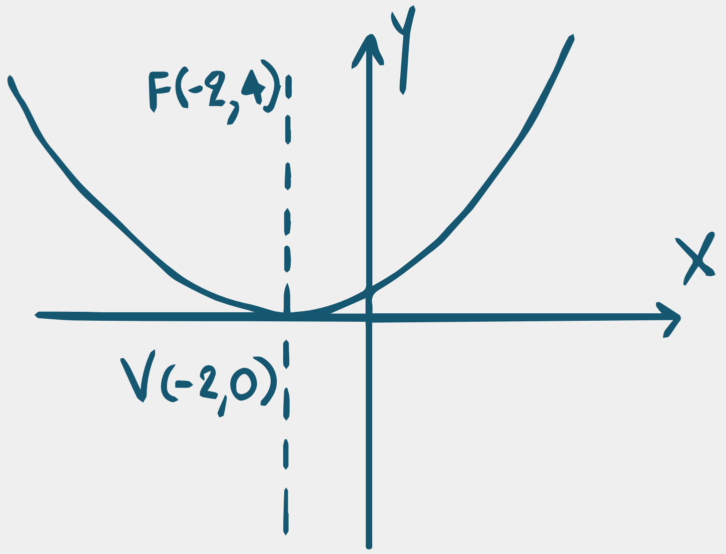
ỨNG DỤNG

Hidden Markov Models (HMMs) trong ML: nhận dạng tiếng nói, xử lý ngôn ngữ tự nhiên

Weather and Climate Modeling

PageRank Algorithm: Google's PageRank algorithm for ranking web pages is based on a Markov chain model of web page navigation.

Stock Price Modeling, Economic Forecasting



THANK YOU!