

Machine Learning

Lecturer: Doctor Bui Thanh Hung

Data Science Laboratory

Faculty of Information Technology

Industrial University of Ho Chi Minh city

Email: hung.buithanhcs@gmail.com (buithanhhung@iuh.edu.vn)

Website: <https://sites.google.com/site/hungthanhbui1980/>

Bài 1:

Viết hàm tính kỳ vọng theo định nghĩa như sau:

- **Kỳ vọng** (*expected value*): là giá trị trung bình sau khi lặp lại một thí nghiệm vô số lần
- Ký hiệu: $E(X)$ hoặc μ
- Công thức:
 - Trường hợp biến ngẫu nhiên rời rạc

$$\mu = E(X) = \sum_x xf(x)$$

- Trường hợp biến ngẫu nhiên liên tục

$$\mu = E(X) = \int_x xf(x)$$

Bài 2:

Viết hàm tính phương sai theo định nghĩa như sau:

- **Phương sai** (*variance*): là trung bình của tổng bình phương độ lệch của tất cả giá trị của biến ngẫu nhiên so với giá trị kỳ vọng
- Ký hiệu: σ^2 , $\text{var}(x)$, $V(x)$
- Công thức:
 - Trường hợp biến ngẫu nhiên rời rạc

$$\sigma^2 = E[(X - \mu)^2] = \sum_x f(x) \times (x - \mu)^2 = E(X^2) - \mu^2 = \sum_x x^2 f(x) - \mu^2$$

- Trường hợp biến ngẫu nhiên liên tục

$$\sigma^2 = E[(X - \mu)^2] = \int_x f(x) \times (x - \mu)^2 = E(X^2) - \mu^2 = \int_x x^2 f(x) - \mu^2$$

Bài 3:

Viết hàm tính độ lệch chuẩn theo định nghĩa như sau:

➤ **Độ lệch chuẩn** (*standard variation*): là căn bậc 2 của giá trị phương sai

Ký hiệu: σ hoặc SD(X)

Bài 4:

a. Một nhóm học sinh có 10 em, trong đó có 3 em học loại giỏi, 4 loại khá, còn lại là trung bình. Từ nhóm đó chọn ngẫu nhiên ra 3 học sinh. Gọi X là số học sinh giỏi trong số học sinh chọn ra

- Lập bảng phân phối xác suất của X, đưa dữ liệu bảng này vào file data.txt

Áp dụng các hàm đã viết ở 1,2,3 tính Kỳ vọng, Phương sai, Độ lệch chuẩn cho dữ liệu biến ngẫu nhiên rời rạc được cung cấp ở file data.txt

b. Áp dụng các hàm đã viết ở 1,2,3 tính Kỳ vọng, Phương sai, Độ lệch chuẩn cho dữ liệu biến ngẫu nhiên liên tục như mô tả sau:

Nhu cầu hàng năm về loại hàng hóa A là **biến ngẫu nhiên liên tục X có hàm mật độ** xác suất như sau: (đơn vị: ngàn sản phẩm)

$$f(x) = \begin{cases} k(30-x) & \text{với } x \in [0,30] \\ 0 & x \notin [0,30] \end{cases}$$

c. Sử dụng thư viện có sẵn tính Kỳ vọng, Phương sai, Độ lệch chuẩn cho dữ liệu ở câu a

d. Sử dụng thư viện có sẵn tính Kỳ vọng, Phương sai, Độ lệch chuẩn cho dữ liệu ở câu b