

# AIO2024 - Module 2

## Module Exam

16th September 2024

- 
- Giờ làm bài từ 6:00 tối thứ 6 (16/08/2024) đến 6:00 sáng thứ 7 (17/08/2024)
  - Khuyến khích các bạn dùng chatbot hay tìm kiếm kết quả trên Internet (miễn các bạn hiểu được đáp án các bạn đang chọn)
  - Contributors: Đăng Nhà, Trường Bình, Anh Khôi, Dương Thuận và Quang Vinh
- 

## 1 Statistics

**Problem 1:** Biến ngẫu nhiên liên tục là gì?

- (a) Một hàm ánh xạ từ không gian mẫu đến tập số thực với miền giá trị hữu hạn
- (b) Một hàm ánh xạ từ không gian mẫu đến tập số thực với miền giá trị không đếm được
- (c) Một hàm ánh xạ từ tập số thực đến không gian mẫu
- (d) Một hàm ánh xạ từ tập số nguyên đến tập số thực

**Problem 2:** Hàm phân phối xác suất (Probability Distribution Function) của một biến ngẫu nhiên  $X$  được định nghĩa là gì?

- (a)  $P(X = x)$
- (b)  $P(X \leq x)$
- (c)  $E[X]$
- (d)  $Var(X)$

**Problem 3:** Công thức tính giá trị kỳ vọng (Expected value) của một biến ngẫu nhiên rời rạc  $X$  là gì?

- (a)  $E[X] = \sum_i^n x \times P(X = x)$
- (b)  $E[X] = \sum_i^n (x - \mu)^2 \times P(X = x)$
- (c)  $E[X] = \sqrt{Var(X)}$
- (d)  $E[X] = \sum_i^n P(X = x)$

**Problem 4:** Công thức tính phương sai (Variance) của một biến ngẫu nhiên  $X$  là gì?

- (a)  $Var(X) = E[X] - \mu^2$

- (b)  $Var(X) = E[X^2] - E[X]^2$
- (c)  $Var(X) = E[(X - \mu)^2]$
- (d)  $Var(X) = \sqrt{E[X^2] - E[X]^2}$

**Problem 5:** Công thức tính độ lệch chuẩn (Standard Deviation) của một biến ngẫu nhiên  $X$  là gì?

- (a)  $\sigma(X) = E[X]$
- (b)  $\sigma(X) = Var(X)$
- (c)  $\sigma(X) = \sqrt{Var(X)}$
- (d)  $\sigma(X) = E[(X - \mu)^2]$

**Problem 6:** Công thức tính hiệp phương sai (Covariance) giữa hai biến ngẫu nhiên  $X$  và  $Y$  là gì?

- (a)  $Cov(X, Y) = E[XY] - E[X]E[Y]$
- (b)  $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$
- (c)  $Cov(X, Y) = E[X]E[Y] - E[XY]$
- (d)  $Cov(X, Y) = \sigma_X * \sigma_Y$

**Problem 7:** Trong một thí nghiệm tung xúc xắc công bằng,  $X$  là số chấm xuất hiện. Tính  $E[X^2]$ .

- (a) 15.17
- (b) 15.33
- (c) 15.50
- (d) 15.67

**Problem 8:** Cho hai biến ngẫu nhiên  $X = \{1, 2, 3, 4\}$  và  $Y = \{2, 3, 4, 3\}$ . Tính hiệp phương sai  $Cov(X, Y)$ .

- (a) 0.5
- (b) 0.6
- (c) 0.7
- (d) 0.8

**Problem 9:** Trong một trò chơi, bạn sẽ mất 1\$ nếu sinh viên kế tiếp tham gia vào lớp AIO2025 thích toán, ngược lại bạn nhận được 2\$ nếu sinh viên ấy không thích toán. Xác suất một sinh viên thích toán là 0.3. Tính giá trị kỳ vọng của trò chơi này.

- (a) -1.1\$
- (b) 1.1\$

(c) 2\$

(d) 4\$

**Problem 10:** Hệ số tương quan Pearson có phạm vi giá trị nào?

(a)  $[0, 1]$

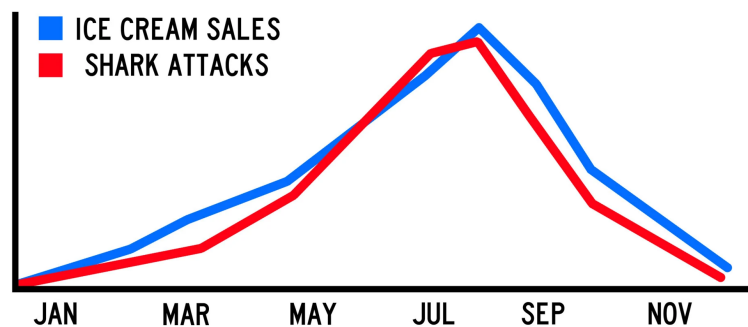
(b)  $[-1, 1]$

(c)  $[-\infty, +\infty]$

(d)  $[0, +\infty]$

**Problem 11:** Câu phát biểu nào đúng nhất cho hình bên dưới?

## CORRELATION IS NOT CAUSATION!



Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)

(a) Hai biến có tương quan luôn có quan hệ nhân quả

(b) Hai biến có tương quan không bao giờ có quan hệ nhân quả

(c) Tương quan và nhân quả là hai khái niệm hoàn toàn độc lập

(d) Tương quan không nhất thiết có quan hệ nhân quả

**Problem 12:** Hãy tìm giá trị "Degrees of Freedom" (bậc tự do) với một mẫu dữ liệu có số lượng mẫu  $N = 30$ ?

(a) 29

(b) 30

(c) 31

(d) 1

**Problem 13:** Cho hai biến ngẫu nhiên  $X$  và  $Y$  có hệ số tương quan là 0.6. Biết rằng  $Var(X) = 4$  và  $Var(Y) = 9$ . Tính  $Cov(X, Y)$  và giải thích ý nghĩa của kết quả này trong việc mô tả mối quan hệ giữa  $X$  và  $Y$ .

- (a)  $Cov(X, Y) = 0.36$ ,  $X$  và  $Y$  có mối quan hệ tuyến tính khá yếu và cùng chiều
- (b)  $Cov(X, Y) = 0.36$ ,  $X$  và  $Y$  có mối quan hệ tuyến tính khá yếu và ngược chiều
- (c)  $Cov(X, Y) = 3.6$ ,  $X$  và  $Y$  có mối quan hệ tuyến tính khá mạnh và ngược chiều
- (d)  $Cov(X, Y) = 3.6$ ,  $X$  và  $Y$  có mối quan hệ tuyến tính khá mạnh và cùng chiều

**Problem 14:** Bạn đang phân tích mối quan hệ giữa nhiệt độ hàng ngày và doanh số bán kem trong một thành phố. Bạn có dữ liệu về nhiệt độ ( $^{\circ}\text{C}$ ) và doanh số bán kem (\$) trong 30 ngày. Hệ số tương quan Pearson giữa hai biến này là 0.75. Nếu bạn chuyển đổi nhiệt độ từ ( $^{\circ}\text{C}$ ) sang ( $^{\circ}\text{F}$ ) và doanh số bán kem từ (\$) sang (€), hệ số tương quan sẽ thay đổi như thế nào? Biết  $^{\circ}\text{F} = ^{\circ}\text{C} \times \frac{9}{5} + 32$  và  $1\$ = 0.91\text{€}$

- (a) Hệ số không đổi
- (b) Hệ số tăng lên
- (c) Hệ số giảm xuống
- (d) Không thể tính hệ số

## 2 Probability

**Problem 15:** Công thức tính xác suất cổ điển là gì?

- (a)  $P(A) = \frac{\text{Số kết quả thuận lợi}}{\text{Tổng số kết quả có thể}}$
- (b)  $P(A) = \frac{\text{Số kết quả không thuận lợi}}{\text{Tổng số kết quả có thể}}$
- (c)  $P(A) = \frac{\text{Tổng số kết quả có thể}}{\text{Số kết quả thuận lợi}}$
- (d)  $P(A) = 1 - \frac{\text{Số kết quả thuận lợi}}{\text{Tổng số kết quả có thể}}$

**Problem 16:** Công thức tính xác suất có điều kiện  $P(A | B)$  là gì?

- (a)  $P(A | B) = \frac{P(A \cap B)}{P(A)}$
- (b)  $P(A | B) = \frac{P(A \cap B)}{P(B)}$
- (c)  $P(A | B) = \frac{P(A)}{P(B)}$
- (d)  $P(A | B) = \frac{P(B)}{P(A \cap B)}$

**Problem 17:** Công thức nào sau đây là công thức cộng xác suất cho hai biến cố tổng quát?

- (a)  $P(A \cup B) = P(A) + P(B)$

(b)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(c)  $P(A \cup B) = P(A) \times P(B)$

(d)  $P(A \cup B) = 1 - P(A' \cap B')$

**Problem 18:** Công thức của quy tắc nhân xác suất (Multiplication Rule) cho hai sự kiện  $A$  và  $B$  là gì?

(a)  $P(A \cap B) = P(A) + P(B)$

(b)  $P(A \cap B) = P(A) \times P(B)$

(c)  $P(A \cap B) = P(A) \times P(B | A)$

(d)  $P(A \cap B) = \frac{P(A)}{P(B)}$

**Problem 19:** Trong một hộp có 10 quả bóng, trong đó có 4 quả màu đỏ và 6 quả màu xanh. Nếu lấy ngẫu nhiên lần lượt 2 quả bóng (mỗi lần 1 quả) không hoàn lại, xác suất để cả hai quả đều màu đỏ là bao nhiêu?

(a)  $\frac{2}{15}$

(b)  $\frac{1}{2}$

(c)  $\frac{3}{25}$

(d)  $\frac{4}{24}$

**Problem 20:** Một công ty sản xuất 1000 sản phẩm mỗi ngày. Biết rằng 2% sản phẩm bị lỗi. Nếu chọn ngẫu nhiên một sản phẩm, xác suất để sản phẩm đó không bị lỗi là bao nhiêu?

(a) 0.02

(b) 0.98

(c) 0.2

(d) 0.8

**Problem 21:** Trong một cuộc khảo sát, 30% người được hỏi thích Pizza, 40% thích Burger, và 15% thích cả hai. Xác suất để một người được chọn ngẫu nhiên không thích cả Pizza lẫn Burger là bao nhiêu?

(a) 0.85

(b) 0.30

(c) 0.58

(d) 0.45

**Problem 22:** Một hệ thống máy tính gồm 3 bộ phận độc lập. Xác suất hoạt động của mỗi bộ phận lần lượt là 0.9, 0.95 và 0.85. Hệ thống hoạt động khi ít nhất 2 trong 3 bộ phận hoạt động. Tính xác suất để hệ thống hoạt động.

(a) 0.892

- (b) 0.974
- (c) 0.247
- (d) 0.727

**Problem 23:** Công thức Bayes được viết như thế nào để hợp lý cho bài toán với  $X$  là các đặc trưng và  $Y$  là lớp?

- (a)  $P(Y | X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$
- (b)  $P(Y | X) = \frac{P(X|Y) \cdot P(X)}{P(Y)}$
- (c)  $P(Y | X) = \frac{P(X|Y)}{P(X) \cdot P(Y)}$
- (d)  $P(Y | X) = P(X | Y) \cdot P(Y) \cdot P(X)$

**Problem 24:** Trong phân loại Naive Bayes, giả định nào được đưa ra về các đặc trưng (features)?

- (a) Các đặc trưng phải phụ thuộc lẫn nhau
- (b) Các đặc trưng phải độc lập có điều kiện với nhau khi biết lớp
- (c) Các đặc trưng phải tuân theo phân phối chuẩn
- (d) Các đặc trưng phải là biến liên tục

**Problem 25:** Trong phân phối chuẩn (Gaussian distribution), tham số  $\mu$  và  $\sigma$  lần lượt đại diện cho:

- (a) Trung bình và phương sai
- (b) Trung vị và độ lệch chuẩn
- (c) Trung bình và độ lệch chuẩn
- (d) Mode và phương sai

**Problem 26:** Cho một tập dữ liệu với 2 lớp và 2 đặc trưng liên tục. Nếu sử dụng Gaussian Naive Bayes để phân loại, cần ước lượng bao nhiêu tham số cho các phân phối trong mô hình?

- (a) 4
- (b) 6
- (c) 8
- (d) 10

**Problem 27:** Khi áp dụng mô hình Gaussian Naive Bayes cho dữ liệu liên tục, tại sao chúng ta lại sử dụng hàm mật độ xác suất thay vì xác suất trực tiếp cho các giá trị của biến?

- (a) Vì dữ liệu liên tục có vô số giá trị có thể
- (b) Vì dữ liệu liên tục luôn tuân theo phân phối chuẩn
- (c) Vì dễ tính toán hơn

(d) Vì cho kết quả chính xác hơn trong mọi trường hợp

**Problem 28:** Cho một tập dữ liệu phân loại hoa Iris với 2 lớp và 2 đặc trưng: chiều dài (Length) và chiều rộng (Width) của cánh hoa. Các thông số thống kê như sau: Với một mẫu mới có Length = 5.5 và Width = 3.0, hãy tính xác suất  $P(\text{Class 1} \mid \text{Length} = 5.5, \text{Width} = 3.0)$  sử dụng Gaussian Naive Bayes.

(a) 0.1357

(b) 0.1753

(c) 0.1537

(d) 0.1573

**Problem 29:** Một bệnh hiểm gắp có tỷ lệ mắc trong dân số là 1%. Một xét nghiệm chẩn đoán bệnh này có độ nhạy (Sensitivity) là 95% (xác suất xét nghiệm dương tính khi người đó thực sự mắc bệnh) và độ đặc hiệu (Specificity) là 90% (xác suất xét nghiệm âm tính khi người đó không mắc bệnh). Nếu một người có kết quả xét nghiệm dương tính, tính xác suất người đó thực sự mắc bệnh, dùng Naive Bayes.

(a) 0.0856

(b) 0.0866

(c) 0.0876

(d) 0.0886

### 3 Bernoulli Naïve Bayes Classifier

Một công ty bảo hiểm muốn dự đoán liệu một khách hàng có khả năng mua bảo hiểm xe hơi hay không dựa trên hai đặc điểm: Tuổi (Trẻ, Trung niên, Già) và Thu nhập (Thấp, Trung bình, Cao). Dữ liệu huấn luyện như sau:

Tuổi	Thu nhập	Mua bảo hiểm
Trẻ	Thấp	Không
Trẻ	Trung bình	Có
Trung niên	Thấp	Có
Già	Trung bình	Có
Già	Cao	Không
Trung niên	Cao	Có
Trẻ	Cao	Không
Trung niên	Trung bình	Có
Già	Thấp	Có
Trẻ	Trung bình	Không

Các câu hỏi sau đây sẽ dựa trên bài toán mô tả ở phần trên. Đọc kỹ đề bài trước khi trả lời các câu hỏi.

**Problem 30:** Dựa vào dữ liệu huấn luyện, xác suất tiên nghiệm (Prior) của việc "Có" mua bảo hiểm là bao nhiêu?

- (a) 0.50
- (b) 0.40
- (c) 0.60
- (d) 0.80

**Problem 31:** Xác suất có điều kiện  $P(\text{Tuổi} = \text{Trung niên} \mid \text{Mua bảo hiểm} = \text{Có})$  là:

- (a) 0.20
- (b) 0.50
- (c) 0.80
- (d) 1.00

**Problem 32:** Xác suất có điều kiện  $P(\text{Thu nhập} = \text{Cao} \mid \text{Mua bảo hiểm} = \text{Có})$  là:

- (a) 0.167
- (b) 0.20
- (c) 0.33
- (d) 0.83

**Problem 33:** Đối với một khách hàng "Trung niên" với Thu nhập "Cao", xác suất mua bảo hiểm là:

- (a)  $P(\text{Có} \mid \text{Trung niên, Cao}) = 0.70$ ,  $P(\text{Không} \mid \text{Trung niên, Cao}) = 0.30$
- (b)  $P(\text{Có} \mid \text{Trung niên, Cao}) = 0.60$ ,  $P(\text{Không} \mid \text{Trung niên, Cao}) = 0.00$
- (c)  $P(\text{Có} \mid \text{Trung niên, Cao}) = 0.42$ ,  $P(\text{Không} \mid \text{Trung niên, Cao}) = 0.58$
- (d)  $P(\text{Có} \mid \text{Trung niên, Cao}) = 1.00$ ,  $P(\text{Không} \mid \text{Trung niên, Cao}) = 0.00$

**Problem 34:** Trong trường hợp này, giả định "Naive" của Naive Bayes có thể dẫn đến sai lệch nào?

- (a) Bỏ qua mối quan hệ tiềm ẩn giữa tuổi và thu nhập
- (b) Đánh giá không chính xác tầm quan trọng của tuổi
- (c) Bỏ qua mối quan hệ tiềm ẩn giữa mua bảo hiểm và thu nhập, tuổi
- (d) Giả định sai về sự phân phối của các đặc trưng

**Problem 35:** Đối với một khách hàng "Già" với Thu nhập "Cao", người ta có mua bảo hiểm hay không:

- (a) Có
- (b) Không
- (c) 50/50



(d) Không biết tính

**Problem 36:** Naive Bayes Classifier được gọi là "Naive" (ngây thơ) vì:

- (a) Nó giả định rằng các đặc trưng (features) độc lập có điều kiện
- (b) Nó giả định rằng các đặc trưng (features) hoàn toàn độc lập
- (c) Nó chỉ xem xét xác suất có điều kiện đơn giản
- (d) Nó bỏ qua mối quan hệ phức tạp giữa các đặc trưng

## 4 Gaussian Naïve Bayes Classifier

Một nhà máy sản xuất bóng đèn LED muốn phân loại sản phẩm thành "Đạt" hoặc "Không đạt" dựa trên hai đặc điểm liên tục: Cường độ ánh sáng (lumen) và Tuổi thọ (giờ). Dưới đây là dữ liệu của 16 bóng đèn LED đã được kiểm tra:

Số thứ tự	Cường độ ánh sáng (lumen)	Tuổi thọ (giờ)	Phân loại
1	805	51000	Đạt
2	780	49500	Đạt
3	795	52000	Đạt
4	820	50500	Đạt
5	810	48000	Đạt
6	785	53000	Đạt
7	815	49000	Đạt
8	800	50000	Đạt
9	610	31000	Không đạt
10	590	28000	Không đạt
11	620	33000	Không đạt
12	580	29000	Không đạt
13	600	32000	Không đạt
14	595	27000	Không đạt
15	605	34000	Không đạt
16	585	31500	Không đạt

Các câu hỏi sau đây sẽ dựa trên bài toán mô tả ở phần trên. Đọc kỹ đề bài trước khi trả lời các câu hỏi.

**Problem 37:** Dựa vào dữ liệu huấn luyện, xác suất tiên nghiệm (prior) của việc bóng đèn "Đạt" là bao nhiêu?

- (a) 0.48
- (b) 0.49
- (c) 0.50
- (d) 0.51

**Problem 38:** Trung bình ( $\mu$ ) của cường độ ánh sáng cho bóng đèn "Đạt" là:

- (a) 801.25 lumen

- (b) 802.50 lumen
- (c) 803.75 lumen
- (d) 805.00 lumen

**Problem 39:** Độ lệch chuẩn ( $\sigma$ ) của tuổi thọ cho bóng đèn "Không đạt" là:

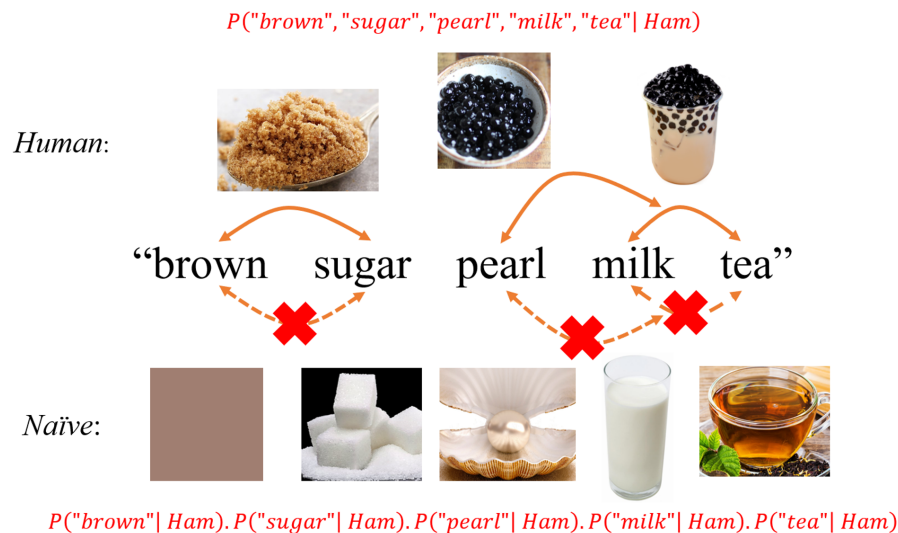
- (a) 2404.04 giờ
- (b) 2304.04 giờ
- (c) 2304.40 giờ
- (d) 2043.04 giờ

**Problem 40:** Xác suất không chuẩn hóa của một bóng đèn có cường độ ánh sáng 750 lumen và tuổi thọ 45000 giờ được phân loại là:

- (a)  $P(\text{Đạt} \mid 750, 45000) \approx 3.75 \times 10^{-11}$ ,  $P(\text{Không đạt} \mid 750, 45000) \approx 1.71 \times 10^{-41}$
- (b)  $P(\text{Đạt} \mid 750, 45000) \approx 6.5 \times 10^{-7}$ ,  $P(\text{Không đạt} \mid 750, 45000) \approx 5.1 \times 10^{-11}$
- (c)  $P(\text{Đạt} \mid 750, 45000) \approx 8.2 \times 10^{-10}$ ,  $P(\text{Không đạt} \mid 750, 45000) \approx 2.6 \times 10^{-6}$
- (d)  $P(\text{Đạt} \mid 750, 45000) \approx 9.25 \times 10^{-11}$ ,  $P(\text{Không đạt} \mid 750, 45000) \approx 7.81 \times 10^{-5}$

## 5 Naive Bayes trong Text Classification

**Problem 41:** Thuật toán Naive Bayes trong bài toán phân loại văn bản, giả định "Naive" là gì?



- (a) Tất cả văn bản đều có xác suất như nhau để thuộc vào 1 lớp
- (b) Độ dài của văn bản quyết định nó thuộc lớp nào

- (c) Sự xuất hiện của một từ không ảnh hưởng đến sự xuất hiện của các từ khác
- (d) Các văn bản của một lớp luôn chứa các từ khóa cụ thể

**Problem 42:** Trong phân loại văn bản sử dụng Naive Bayes, tại sao chúng ta thường sử dụng Log Likelihood thay vì xác suất thông thường?

- (a) Để tăng độ chính xác của mô hình
- (b) Để giảm thời gian tính toán
- (c) Để tránh Underflow khi nhân nhiều số nhỏ
- (d) Để đơn giản hóa quá trình huấn luyện mô hình

**Problem 43:** Trong kỹ thuật Laplacian Smoothing, tại sao chúng ta thêm 1 vào tử số và mẫu số khi tính xác suất của một từ?

- (a) Để tăng độ chính xác của mô hình
- (b) Để giảm thời gian tính toán
- (c) Để chuẩn hóa dữ liệu
- (d) Để tránh xác suất bằng 0 cho các từ chưa xuất hiện

**Problem 44:** Chọn đáp án đúng.

Một hàm Python thực hiện Laplacian Smoothing cho một từ điển tần suất từ. Hàm này nhận vào hai tham số:

- Một từ điển `word_freq` có dạng `{word: frequency}`, trong đó `frequency` là số lần xuất hiện của từ `word` trong dữ liệu.
- Tổng số từ `total_words` trong dữ liệu.

Hàm cần trả về một từ điển mới với xác suất của các từ đã được điều chỉnh bằng phương pháp Laplacian Smoothing.

(a)

```
1 def laplacian_smoothing(word_freq, total_words):  
2     return {word: (freq + 1) / (total_words + 1) for word, freq in word_freq.items()}
```

(b)

```
1 def laplacian_smoothing(word_freq, total_words):  
2     return {word: (freq + 1) / (total_words + len(word_freq)) for word, freq in  
           word_freq.items()}
```

(c)

```
1 def laplacian_smoothing(word_freq, total_words):  
2     return {word: freq / total_words for word, freq in word_freq.items()}
```

(d)

```
1 def laplacian_smoothing(word_freq, total_words):  
2     vocab_size = len(word_freq)  
3     return {word: freq / (total_words + vocab_size) for word, freq in word_freq.items()}
```

## 6 Coding (ưu tiên việc tìm kiếm và dùng các hàm cho sẵn trong python)

**Problem 45:** Hoàn thiện đoạn code và xác định đâu là test cases đúng cho bài toán sau:

### Mean

- Data:  $X = x_1, \dots, x_N$
- Mean:  $\mu = \frac{1}{N} \sum_{i=1}^N x_i$

Cho tập data  $X$ . Tìm giá trị trung bình (mean) của data  $X$ .

```
1
2 def calculate_mean(self, X):
3     # Your code here
4     pass
```

(a)  $X = [-2, 7, 3, 0, 6, 1, -3, 6, 5, -3]$ , Mean = 2.3

(b)  $X = [10, 20, 30, 40, 50]$ , Mean = 30.0

(c)  $X = [0, 0, 0, 0, 0]$ , Mean = 0.0

(d)  $X = [-1, -1, -1, -1, -1]$ , Mean = -1.0

**Problem 46:** Hoàn thiện đoạn code và xác định đâu là test cases đúng cho bài toán sau:

### Median

- Data:  $X = x_1, \dots, x_N$
- Median:
  - Sort  $X \rightarrow S$
  - If  $N$  is odd  $m = \frac{S_{N+1}}{2}$
  - If  $N$  is even  $m = \frac{1}{2}(S_{\frac{N}{2}} + S_{\frac{N}{2}+1})$

Do tập data  $X$ . Tìm trung vị (median) của  $X$ .

```
1
2 def calculate_median(self, X):
3     # Your code here
4     pass
```

- (a)  $X = [7, 0, 9, 2, 5, 7]$ , Median = 5.5  
 (b)  $X = [1, 3, 5, 7, 9]$ , Median = 6.0  
 (c)  $X = [2, 5, 7, 0]$ , Median = 5.0  
 (d)  $X = [-1, 4, 7, 9, 10]$ , Median = 7.0

**Problem 47:** Hoàn thiện đoạn code và xác định đâu là test cases đúng cho bài toán sau:

### Variance

- Data:  $X = x_1, \dots, x_N$
- Mean:  $\mu = \frac{1}{N} \sum_{i=1}^N x_i$
- Variance:  $var(X) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$
- Standard Deviation:  $\sigma = \sqrt{var(X)}$

Cho tập data X, tính phương sai (variance), độ lệch chuẩn (standard deviation) của X.

```

1
2  def calculate_variance_std(self, data)
3      # Your code here
4      # variance = ...
5      # std_dev = ...
6      pass

```

- (a)  $X = [18, 92, 5, 34, 3, 5, 70, 63, 8, 28, 38, 47, 85, 68, 23]$ , (variance, std\_dev) = (851.9822222222222, 29.188734508748784)  
 (b)  $X = [84, 30, 26, 85, 40, 69, 28, 28, 85, 33, 81, 29, 9, 24, 3]$ , (variance, std\_dev) = (756.9066666666666, 27.87304552191358)  
 (c)  $X = [56, 78, 73, 32, 24, 11, 96, 34, 57, 13, 65, 16, 92, 4, 29]$ , (variance, std\_dev) = (669.0222222222221, 29.479182862186363)  
 (d)  $X = [41, 35, 89, 1, 35, 90, 58, 61, 99, 25, 75, 9, 30, 6, 44]$ , (variance, std\_dev) = (827.4488888888889, 30.28941876115963)

**Problem 48:** Hoàn thiện đoạn code và xác định đâu là test cases đúng cho bài toán sau:

### Correlation Coefficient

- Data(Random variables X, Y):  $X = \{x_1, \dots, x_N\}$   $Y = \{y_1, \dots, y_N\}$
- Correlation Coefficient: 
$$p_{xy} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \frac{n(\sum_i x_i y_i) - (\sum_i x_i)(\sum_i y_i)}{\sqrt{n \sum_i x_i^2 - (\sum_i x_i)^2} \sqrt{n \sum_i y_i^2 - (\sum_i y_i)^2}}$$

Cho hai vector X và Y. Tính độ tương quan giữa 2 vector.

```

1
2 def calculate_correlation_coefficient(self, X, Y):
3     # Your code here
4     pass

```

- (a) X1: [50 53 97 45 81 61 87 66 9 34], Y1: [18 46 4 54 40 32 36 2 25 19], cor = - 0.75
- (b) X2: [13 42 17 2 83 57 19 54 19 10], Y2: [51 85 2 97 86 60 76 83 17 61], cor = 0.5
- (c) X3: [27 38 93 60 23 10 53 46 94 3], Y3: [ 6 70 3 65 61 76 15 9 4 25], cor = 0.32
- (d) X4: [23 45 77 20 89 83 87 75 56 12], Y4: [13 53 62 54 57 27 29 46 79 5], cor = -0.15

**Problem 49:** Hoàn thiện đoạn code và xác định đâu là test cases đúng cho bài toán sau:

### Dot product

- Vector:  $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix}$   $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix}$
- Dot Product:  $\mathbf{v} \cdot \mathbf{u} = v_1 * u_1 + v_2 * u_2 + \dots + v_n * u_n$

Cho hai vector u và v. Tính tích vô hướng u . v.

```

1
2 def calculate_dot_product(self, v, u):
3     # Your code here
4     pass

```

- (a)  $\mathbf{v} = [1, 2, 3, 4]$ ,  $\mathbf{u} = [3, 4, 5, 6]$ ,  $\mathbf{v} \cdot \mathbf{u} = 40$
- (b)  $\mathbf{v} = [5, 6, 7]$ ,  $\mathbf{u} = [2, 3]$ ,  $\mathbf{v} \cdot \mathbf{u} = 38$
- (c)  $\mathbf{v} = [-1, 2, -3]$ ,  $\mathbf{u} = [4, -5, 6]$ ,  $\mathbf{v} \cdot \mathbf{u} = -32$
- (d)  $\mathbf{v} = [4, 2, 3]$ ,  $\mathbf{u} = [4, 5, 6]$ ,  $\mathbf{v} \cdot \mathbf{u} = 36$

**Problem 50:** Hoàn thiện đoạn code và xác định đâu là test cases đúng cho bài toán sau:

### Multiplying a matrix by a matrix

- Matrix A:  $\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}, \mathbf{A} \in R^{m \times n}$
- Matrix B:  $\mathbf{B} = \begin{bmatrix} b_{11} & \dots & b_{1k} \\ \dots & \dots & \dots \\ b_{n1} & \dots & b_{nk} \end{bmatrix}, \mathbf{B} \in R^{n \times k}$
- $\mathbf{C} = \mathbf{AB} = \begin{bmatrix} a_{11} * b_{11} + \dots + a_{1n} * b_{n1} & \dots & a_{11} * b_{1k} + a_{1n} * b_{nk} \\ \dots & \dots & \dots \\ a_{m1} * b_{11} + \dots + a_{mn} * b_{n1} & \dots & a_{m1} * b_{1k} + a_{mn} * b_{nk} \end{bmatrix},$   
 $\mathbf{C} \in R^{m \times k}$

Cho hai ma trận A và B. Ma trận C là kết quả nhân ma trận (Matrix multiplication) của ma trận A và B. Tính C.

```

1
2 def calculate_matrix_product(self, A, B):
3     # Your code here
4     pass

```

- (a)  $\mathbf{A} = \begin{bmatrix} 8 & 8 & 1 \\ 2 & 7 & 1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 8 & 1 \\ 7 & 1 \\ 2 & 8 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 122 & 24 \\ 65 & 9 \end{bmatrix}$
- (b)  $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$
- (c)  $\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 3 & 4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \\ 3 & 4 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 2 & 5 & 9 \\ 3 & 4 & 12 \end{bmatrix}$
- (d)  $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix}$

**Problem 51:** Hoàn thiện đoạn code và xác định đâu là test cases đúng cho bài toán sau:

### Eigenvector and eigenvalue

- $\mathbf{A} \in R^{n \times n}, \mathbf{I}(\text{identity matrix}) \in R^{n \times n}, \mathbf{v} \in R^n$
- Eigenvalue ( $\lambda$ ):  $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$
- Eigenvector ( $\mathbf{v}$ ):  $\mathbf{Av} = \lambda \mathbf{v} \iff \mathbf{Av} - \lambda \mathbf{Iv} =$
- Normalize vector:  $\frac{\mathbf{v}}{\|\mathbf{v}\|}, v_i = \frac{v_i}{\sum_1^n v_i^2}$

Cho vector A. Tìm Eigenvector ( $\mathbf{v}$ ) đã được normalize và eigenvalue  $\lambda$  của A.

```

1
2 def calculate_eigenvalues_eigenvectors(self, A):
3     def get_eigenvalues(matrix):

```

```
4      # Your code here
5      def get_eigenvectors(matrix, eigenvalue):
6          # Your code here
7      pass
```

(a)  $A = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$ ,  $\{\lambda_1 : \mathbf{v}_1, \lambda_2 : \mathbf{v}_2\} = \{4.0: [-0.7071067811865476, -0.7071067811865476], -3.0: [-0.6, 0.8]\}$

(b)  $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ ,  $\{\lambda_1 : \mathbf{v}_1, \lambda_2 : \mathbf{v}_2\} = \{3.0: [-0.7071067811865475, -0.7071067811865475], 1.0: [-0.7071067811865475, 0.7071067811865475]\}$

(c)  $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ ,  $\{\lambda_1 : \mathbf{v}_1, \lambda_2 : \mathbf{v}_2\} = \{1.0: [0.7071067811865475, 0.7071067811865475], -1.0: [-0.7071067811865475, 0.7071067811865475]\}$

(d)  $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ ,  $\{\lambda_1 : \mathbf{v}_1, \lambda_2 : \mathbf{v}_2\} = \{6.464101615137754: [-0.9390708015880442, -0.3437237693334403], 0.4641016151377544: [0.5906904945688722, 0.8068982213550735]\}$

---

**Hết**

---