Machine Learning exam

I. Lý thuyết.

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$$

1) Vector $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ và $\vec{v}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ độc lập tuyến tính, vì với mọi số thực $c_1, c_2 \in \mathbb{R}$ bất kỳ:

$$c_1 v_1 + c_2 v_2 = A\vec{c} = 0$$

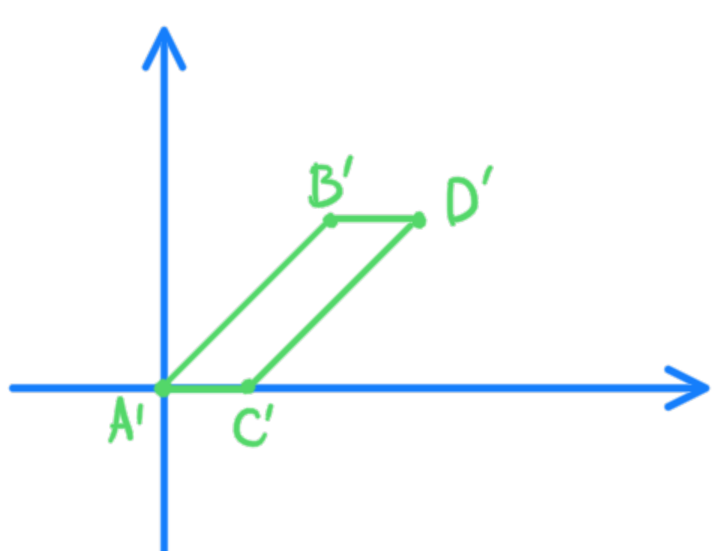
có nghiệm $\vec{c} \neq \vec{0}$ khi và chỉ khi A là ma trận singular.
Mà $\det(A) = 2$, do đó A là ma trận khả nghịch.

↳ Chỉ có $\vec{c} = \vec{0}$ thỏa mãn $A\vec{c} = 0$. Như vậy $\vec{v}_1$ và $\vec{v}_2$ độc lập tuyến tính. □

2) Biến đổi mỗi điểm với ma trận A:

$$A\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$A\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$



↳ Linear transformation với ma trận A cho hình mới A'B'D'C' với diện tích là 2.

3) $A = PDP^{-1}$, trong đó $\begin{cases} P: \text{ ma trận eigenvectors của } A. \\ D: \text{ ma trận eigenvalues của } A. \end{cases}$

☆ Tìm eigenvalues λ_i của A:

$$A\vec{x} = \lambda\vec{x} \Leftrightarrow (A-\lambda I)\vec{x} = 0$$

Ma trận A có eigenvector $\vec{x} \neq \vec{0}$ khi và chỉ khi $A-\lambda I$ là ma trận singular.

$\Leftrightarrow \det(A-\lambda I) = 0$

$\Leftrightarrow \begin{vmatrix} 1-\lambda & 2 \\ 0 & 2-\lambda \end{vmatrix} = 0$

$\Leftrightarrow (1-\lambda)(2-\lambda) = 0$

$\Leftrightarrow \begin{cases} \lambda_1 = 1 \\ \lambda_2 = 2 \end{cases}$. Từ đó: $D = \begin{bmatrix} 1 & \\ & 2 \end{bmatrix}$.

☆ Tìm eigenvectors của A:

+) $\lambda_1 = 1$: $\begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \vec{x}_1 \end{bmatrix} = 0 \Leftrightarrow \vec{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

+) $\lambda_2 = 2$: $\begin{bmatrix} -1 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \vec{x}_2 \end{bmatrix} = 0 \Leftrightarrow \vec{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

Từ đó: $P = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. Suy ra $P^{-1} = \frac{1}{\det P} C^T = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$

(C là cofactor matrix của P: $C = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$)

☆ Đáp án:

$$\boxed{\underbrace{\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}}_{P}\underbrace{\begin{bmatrix} 1 & \\ & 2 \end{bmatrix}}_{D}\underbrace{\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}}_{P^{-1}}}$$

4) Ta có $A^4 = PDP^{-1}PDP^{-1}PDP^{-1}PDP^{-1}$

+) Với mỗi $P^{-1}P = I$, ta rút gọn A^4 thành:

$A^4 = PD^4P^{-1}$

(từ đó có công thức $A^k = PD^kP^{-1}$)

$$A^4 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & \\ & 16 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

$$A^4 = \begin{bmatrix} 1 & 32 \\ 0 & 16 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

$A^4 = \begin{bmatrix} 1 & 30 \\ \dots \end{bmatrix}$

$\begin{bmatrix} 0 & 16 \end{bmatrix}$

5) Sử dụng tính chất của determinant:

(a) $\det(AB) = \det A \det B$

(b) $\det A^{-1} = \frac{1}{\det A}$.

(c) Determinant của ma trận chéo là tích của mỗi entry trên đường chéo đó.

+) Ta có ma trận A khả nghịch $\Rightarrow \det A \neq 0$ và do đó $\lambda_i \neq 0$. Như vậy A có thể chéo hóa:

$$A = PDP^{-1}$$

Từ tính chất (a)(b)(c) ta có:

$$\det A = \det P \det D \det P^{-1}$$

$$\Rightarrow \det A = \det P \left(\prod_{i=1}^{n} \lambda_i\right) \frac{1}{\det P}$$

$$\Rightarrow \boxed{\det A = \prod_{i=1}^{n} \lambda_i} \quad \square$$

6) $f(x) = x \log(x)$ — giả định là $\log_e(x) = \ln(x)$.

+) Đạo hàm của $f(x)$:

$$\frac{df}{dx} = \ln(x) + x \cdot \frac{1}{x} = \ln(x) + 1.$$

+) Min $f(x)$ xảy ra ở $\frac{df}{dx} = 0$:

$$\ln(x) + 1 = 0$$

$$\ln(x) = -1$$

$$x = e^{-1}.$$

+) Thay $x = e^{-1}$ vào $f(x) = x \ln(x)$:

$$f(e^{-1}) = e^{-1} \ln(e^{-1}) = -e^{-1}.$$

Đáp số: $\boxed{f(x) \text{ đạt min tại } (x, f(x)) = (e^{-1}, -e^{-1})}$

7) Taylor expansion của f(x):

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i!}(x-a)^i;$$

$f^{(i)}(a)$: đạo hàm thứ i của f(x) tại a.

↳ Taylor expansion của $x\ln(x)$ tại $x=1$ tới số hạng x^5 là:

$$f(x) = f(1) + f'(1)(x-1) + \frac{f^{(2)}(1)}{2!}(x-1)^2 + \dots + \frac{f^{(5)}(1)}{5!}(x-1)^5$$

$$= \boxed{\ln 1 + (\ln(1)+1)(x-1) + \frac{1}{2!}(x-1)^2 - \frac{1}{3!}(x-1)^3 + \frac{2}{4!}(x-1)^4 - \frac{6}{5!}(x-1)^5}$$

8) $$L(w) = \|Aw - y\|_2^2 + \lambda\|w\|_2^2$$

$$\begin{cases} A: (m \times n) \text{ matrix} \\ w: (n \times 1) \text{ vector} \\ y: (m \times 1) \text{ vector} \\ \lambda > 0: \text{số vô hướng} \end{cases}$$

+) Đặt ∇_w là gradient của $L(w)$. Định nghĩa của gradient vector:

$$\underset{(n\times1)}{\nabla_w L(w)} = \frac{\partial L(w)}{\partial w} = \left[\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \dots, \frac{\partial L}{\partial w_n}\right]^T$$

+) Ta có gradient của L(w) như sau:

$$\underbrace{\nabla_w L(w)}_{(n\times1)} = \underbrace{2A^T}_{(n\times m)}\underbrace{(Aw-y)}_{(m\times1)} + \underbrace{2\lambda w}_{(n\times1)}$$

+) Thuật toán gradient descent:

- Đặt giá trị khởi tạo w_0.
- Đặt giá trị learning rate η
- Đưa ra dự đoán $\hat{y} = Aw_0$ với toàn bộ ma trận dữ liệu A.
- Tính gradient của hàm $L(w_0)$: $\nabla_{w_0} L(w_0)$
- Cập nhật w mới bằng cách đi ngược chiều của gradient:

$$w_1 = w_0 - \eta \nabla_{w_0} L(w_0)$$

- Lặp lại với một số lượng iteration nhất định hoặc đến khi converge tại min $L(w)$.

☆ Thử các learning rate khác nhau, vì:
  + Quá nhỏ → converge chậm
  + Quá lớn → không converge tại min $L(w)$.

9) $L(w)$ đạt tối ưu khi $\nabla_w L(w) = 0$:

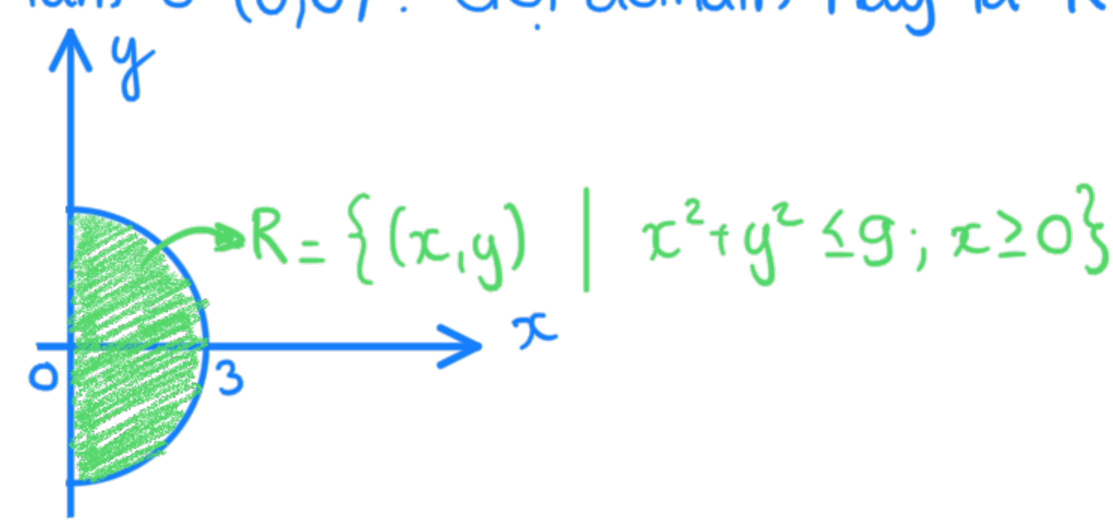
$$2A^TAw - 2A^Ty + 2\lambda w = 0$$

$$(A^TA + \lambda I)\, w = A^Ty$$

$$\boxed{w = (A^TA + \lambda I)^{-1} A^Ty}$$

10) $$I = \iint_{x^2+y^2 \le 9,\ x \ge 0} \sqrt{x^2+y^2}\, dx\, dy$$

+) Domain tích phân: nửa bên phải của hình tròn $x^2+y^2 = 9$ với bán kính 3 và tâm ở $(0,0)$. Gọi domain này là R:



+) Chuyển đổi I sang hệ toạ độ polar (r, θ):

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ dx\,dy = r\,dr\,d\theta \end{cases} \qquad \Rightarrow x^2+y^2 = r^2 \quad (0 < r \le 3)$$

$$\Rightarrow \sqrt{x^2+y^2} = r.$$

$$I = \iint_R r \cdot r\,dr\,d\theta$$

$$= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^3 r^2\, dr\, d\theta$$

$$= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{r^3}{3}\Big|_0^3 \, d\theta$$

$$= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 9 \, d\theta$$

$$= 9\theta \Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}$$

$$= \boxed{9\pi.}$$

II. Thực hành

3) Nhận xét:

+) $g(x)$ là hàm số lẻ : $g(-x) = -g(x)$.

+) Domain của $g(x)$: $(-\infty, \infty)$. Hàm $g(x)$ liên tục và có đạo hàm với mọi x.

+) $$\begin{cases} \lim\limits_{x \to +\infty} g(x) = 1 \\ \lim\limits_{x \to -\infty} g(x) = 0 \\ g(0) = \frac{1}{2} \end{cases}$$

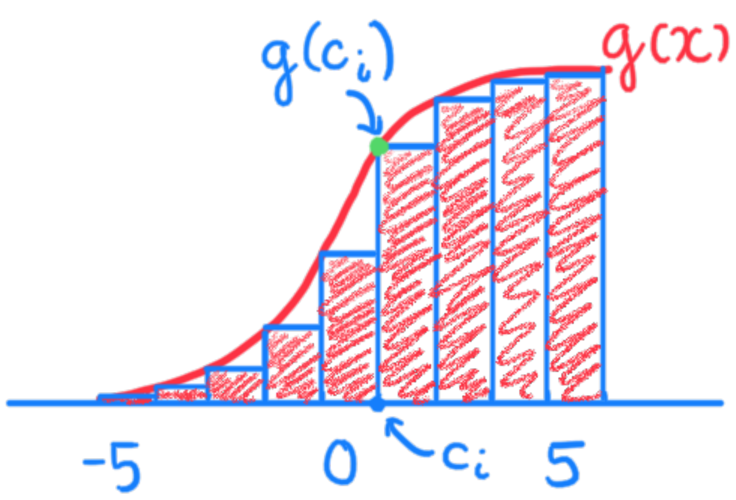
+) $g(x)$ là hàm không tuyến tính, thường sử dụng trong binary classification với mục đích phân loại dữ liệu dựa trên xác suất rơi vào 1 trong 2 class.
(có thể dùng cho multiclass, tuy nhiên không tốt bằng softmax).

5) Sử dụng phương pháp Left Riemann sum:

+) Chia domain (a, b) thành n phần bằng nhau.

+) Vẽ các hình chữ nhật với đáy $\frac{b-a}{n}$ và cạnh bên trái là giá trị của $g(c_i)$ tại điểm c_i. Cạnh bên phải bằng cạnh bên trái.

+) Tổng diện tích của các hình chữ nhật là tích phân của $g(x)$ từ a đến b:



$$\int_a^b g(x)\,dx = \sum_{i=1}^{n} \left(\frac{b-a}{n}\right) g(c_i)$$