BÀI THI GIỮA KÌ

MTH56-2425

Nguyễn Thị Mỹ Kim
MSSV: 23122040

Ngày 25 tháng 4 năm 2025

1

Đặt ngày 14/04 là ngày 0, ngày 15/05 là ngày 1, tương tự như thế ta có ngày 18/4 là ngày 4. Tương ứng N = 4.
Như thế, ta gọi x_n là giá đóng cổ phiếu ngày n, x_{n-1} là giá đóng cổ phiếu của 1 ngày trước ngày n, x_{n-2} là giá đóng cổ phiếu của 2 ngày trước ngày n.

1.a

Ta có giả thiết: Giá đóng cổ phiếu Nvidia là tổ hợp tuyến tính với hệ số thực của 2 ngày trước đó. Như vậy, gọi x_n là giá đóng cổ phiếu ngàu thứ n thì

$$x_n = ax_{n-2} + bx_{n-1}$$

với $a, b \in \mathbb{R}, n \in \overline{2, 4}$
Như thế ta có được gia đình hàm $\mathcal{H} : \{f : \mathbb{R}^2 \to \mathbb{R} \mid f(x_{n-2}, x_{n-1}) = ax_{n-2} + bx_{n-1},\ a, b \in \mathbb{R}\}$

1.b

Ta cần chứng minh tập các hàm tuyến tính hai ẩn:

$$\mathcal{H} : \{f : \mathbb{R}^2 \to \mathbb{R} \mid f(x_{n-2}, x_{n-1}) = ax_{n-2} + bx_{n-1},\ a, b \in \mathbb{R}\}$$

với phép cộng hàm số tuyến tính 2 ẩn và phép nhân hàm số tuyến tính với 2 số thực là không gian vector trên $\mathbb{R}$ với số chiều là 2.
Trước tiên ta chỉ ra $\mathcal{H}$ là một không gian vector.
Với $f_1(x_{n-2}, x_{n-1}) = ax_{n-2} + bx_{n1},\ f_2(x_{n-2}, x_{n-1}) = a_2x_{n-2} + b_2x_{n-1} \in \mathcal{H}$ với $a_1, b_1, a_2, b_2 \in \mathbb{R}$ và $\lambda \in \mathbb{R}$:

- $0 \in \mathcal{H}$ vì $0x_{n-2} + 0x_{n-1} = 0$

- $(f_1 + f_2)(x_{n-2}, x_{n-1}) = (a_1 + a_2)x_{n-2} + (b_1 + b_2)x_{n-1} \in \mathcal{H}$ vì $a_1 + a_2, b_1 + b_2 \in \mathbb{R}$
- $(\lambda f_1)(x_{n-2}, x_{n-1}) = (\lambda a_1)x_{n-2} + (\lambda b_1)x_{n-1} \in \mathcal{H}$ vì $\lambda a_1, \lambda b_1 \in \mathbb{R}$

Vậy $\mathcal{H}$ là một không gian vector trên R.
Tiếp theo, ta chỉ ra $\mathcal{H}$ là không gian vector có số chiều là 2.
Đặt $g_1(x_{n-2}, x_{n-1}) = x_{n-2}$, $g_2(x_{n-2}, x_{n-1}) = x_{n-1}$. Khi đó $\mathcal{H}$ là tỗ hợp tuyến tính của $T = \{g_1, g_2\}$ với T là tập con khác rỗng của $\mathbb{R}$, như vậy $\mathcal{H}$ còn được gọi là không gian sinh bởi T:

$$\mathcal{H} = \langle\{g_1, g_2\}\rangle = \{ag_1 + bg_2,\ a, b \in \mathbb{R}\} = \{ax_{n-2} + bx_{n-1},\ a, b \in \mathbb{R}\}$$

Ta xét:

$$\alpha_1 g_1 + \alpha_2 g_2 = 0 \Leftrightarrow \alpha_1 x_{n-2} + \alpha_1 x_{n-1} = 0 \Leftrightarrow \alpha_1 = \alpha_2 = 0$$

Vậy g_1, g_2 độc lập tuyến tính nên $T = \{g_1, g_2\}$ là một cơ sở của $\mathcal{H}$, suy ra $\dim \mathcal{H} = 2$
Vậy $\mathcal{H}$ là không gian vector trên R và có số chiều là 2

1.c

Tập dữ liệu θ có dạng $\{(x_{n-2}, x_{n-1}, x_n) : x_n \in \mathbb{R}\}$ với $n \in \overline{2, N}$ tương ứng với $n \in \{2, 3, 4\}$ do $N = 4$.
Vậy tập hợp θ gồm 3 điểm dữ liệu trong $\mathbb{R}^3$ cho mô hình hồi qui tuyến tính ở bài 1.a là

$$\begin{aligned}\theta &= \{(x_0, x_1, x_2), (x_1, x_2, x_3), (x_2, x_3, x_4)\} \\ &= \{(109, 105, 95), (105, 95, 96), (95, 96, 98)\}\end{aligned}$$

1.d

Bài toán học từ dữ liệu, cho bộ dữ liệu $\theta = \{(x_0, x_1, x_2), (x_1, x_2, x_3), (x_2, x_3, x_4)\} \subset \mathbb{R}^3$. Tìm đường thẳng $\hat{x}_n = f(x_{n-2}, x_{n-1}) = ax_{n-2} + bx_{n-1}$ tương ứng với việc tìm $f \in H$ sao cho hàm mất mát:

$$\begin{aligned}L(f, \theta) &= \frac{1}{3}\sum_{i=2}^{4}(\hat{x}_n - x_n)^2 \\ &= \frac{1}{3}\sum_{i=2}^{4}(ax_{n-2} + bx_{n-1} - x_n)^2\end{aligned}$$

đạt giá trị nhỏ nhất.
Ta có: x_n phụ phuộc vào x_{n-2} và x_{n-1} tương ứng giá cổ cổ phiếu ngày n phụ thuộc vào giá cổ phiếu ngày $n-2$ và ngày $n-1$. Khi đó, ta sử dụng hàm $\hat{x}_n = ax_{n-2} + bx_{n-1}$ để mô tả qui luật này. Hơn nữa, mô hình này là hồi quy tuyến tính sử dụng các giá trị $(x_{n-2}, x_{n-1}) \in \mathbb{R}^2$ để ánh xạ vào tập $\mathbb{R}$ có được x_n. Các giá trị x_n đã đươc cho trước trong quá trình học. Vì thế đây là bài toán học giám sát.

1.e

Đạo hàm riêng của L theo a:

$$
\begin{aligned}
L_a = \frac{\partial L}{\partial a} &= \frac{1}{3}\sum_{i=2}^{4} 2(ax_{n-2} + bx_{n-1} - x_n)x_{n-2} \\
&= \frac{1}{3}2[(109a + 105b - 95)109 + (105a + 05b - 96)105 + (95a + 96b - 98)95] \\
&= \frac{1}{3}2(31931a + 30540b - 29745)
\end{aligned}
$$

Đạo hàm riêng của L theo b:

$$
\begin{aligned}
L_b = \frac{\partial L}{\partial b} &= \frac{1}{3}\sum_{i=2}^{4} 2(ax_{n-2} + bx_{n-1} - x_n)x_{n-1} \\
&= \frac{1}{3}2(30540a + 29266b - 28503)
\end{aligned}
$$

Ta tìm điểm dừng bằng cách giải hệ:

$$
\begin{cases} \frac{\partial L}{\partial a} = 0 \\ \frac{\partial L}{\partial b} = 0 \end{cases}
$$

$$
\Leftrightarrow \begin{cases} a = 0.01973852972 \\ b = 0.95333100880 \end{cases}
$$

Theo định nghĩa, tập $C \subseteq \mathbb{R}^2$ được gọi là lồi nếu $\forall \vec{x}, \vec{y} \in C$ và $t \in [0, 1]$ thì

$$
(1 - t)\vec{x} + t\vec{y} \in C
$$

Thật vậy, rõ ràng tổ hợp tuyến tính $\vec{z} = (1 - t)\vec{x} + t\vec{y}$ cũng là một vector trong $\mathbb{R}^2$ (phép nhân và cộng trong không gian vector $\mathbb{R}^2$). Vậy đoạn thẳng nối $\vec{x}, \vec{y}$ hoàn toàn nằm trong $\mathbb{R}^2$. Do đó $\mathbb{R}^2$ lồi. Ta lại có:

$$
\begin{aligned}
L_{aa} &= \frac{2}{3} \times 31931 \\
L_{bb} &= \frac{2}{3} \times 29266 \\
L_{ab} = L_{ba} &= \frac{2}{3} \times 30540
\end{aligned}
$$

Do ma trận Hessian $\nabla^2 L(a, b) \succ 0$ và $\operatorname{dom} L = \mathbb{R}^2$ là lồi nên hàm L lồi.
Vậy $(a, b) = (0.01973852972, 0.95333100880)$ là cực tiểu toàn cục của hàm mất mát L hay nghiệm tối ưu của bài toán.

$$
\hat{x}_n = 0.01973852972 \times x_{n-2} + 0.95333100880 \times x_{n-1}
$$

Lớp	Số node	Hàm kích hoạt
Input	2	Không có
Output	1	$g(x) = x$

Bảng 1: Bảng mô tả cấu trúc mạng neuron (câu 1.f)

1.f

Mô tả:

- Hàm không có bias.
- Lớp input bao gồm 2 node tương ứng là x_{n-2} và x_{n-1}. Hàm input không có hàm kích hoạt
- Lớp output là 1 node được kết nối đầy đủ với 2 node của input trước đó. Hàm kích hoạt ở lớp này là $g(x) = x$.
- 2 kết nối đi từ input x_{n-2} và x_{n-1} đến output mang trọng số của mô hình w_1, w_2 (hay ta còn gọi là a và b trong bài toán này).

Nhờ vào cấu trục mạng như trên, input x_{n-2} và x_{n-1} đi vào lớp input được tính theo hàm tuyến tính $z = f(x_{n-2}, x_{n-1}) = w_1x_{n-2} + w_2x_{n-1}$. Giá trị z này đi qua hàm kích hoạt ở lớp output $g(z) = z$. Mô hình sẽ sử dụng hàm loss và thuật toán dùng cho quá trình huấn luyện để tìm ra trọng số w_1, w_2 tương ứng cho từng input là x_{n-2} và x_{n-1}

1.g

Ta có:

$$\begin{aligned}\nabla L &= (\frac{\partial L}{\partial a}, \frac{\partial L}{\partial b})^\top \\ &= [\frac{2}{3}(31931a + 30540b - 29745, \frac{2}{3}(30540a + 29266b - 28503)]^\top\end{aligned}$$

Ta có $\eta = 0.01$, điểm khởi đầu $X_0 = (0.5, 0.5)^\top$

$$X_n = X_{n-1} - \eta\nabla L(X_{n-1)}$$

3 bước đầu tiên của thuật toán gradient descent là

n	X_n	$\nabla L(X_n)$
0	$(0.5,\ 0.5)^\top$	$(993.6666,\ 933.3333)^\top$
1	$(-9.4366,\ 8.8333)^\top$	$(-400558.13555,\ -383476.75555)^\top$
2	$(3996.144688,\ 3825.934222)^\top$	$(162943454.80505183,\ 155989031.16414815)^\top$
3	$(-1625438.4033616,\ -1556064.37741)^\top$	$(-66282739659.41624,\ -63453798274.144135)^\top$

2

2.a

Lập ma trận A từ bảng chỉ số sức khỏe (giả định) của bệnh nhân tiểu đường

$$A = \begin{bmatrix} 4 & 3 & a \\ 2 & b & c \\ 5 & 4 & d \end{bmatrix}$$

Đặt $A = \begin{bmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vec{v}_3 \end{bmatrix}$, với $\vec{v}_i$ là các vector hàng.

$$\vec{v}_1 = (4, 3, a)$$
$$\vec{v}_2 = (2, b, c)$$
$$\vec{v}_3 = (5, 4, d)$$

Giả sử hạng nhỏ nhất là 1, khi đó:

- Ta tìm dạng biểu diễn $\vec{v}_2$ là tổ hợp tuyến tính của $\vec{v}_1$ để:

  $$k\vec{v}_1 = \vec{v}_2 \Leftrightarrow k(4, 3, a) = (2, b, c)$$

  Ta có thể chọn $k = \frac{1}{2}$ do đó $b = \frac{3}{2}, c = \frac{a}{2}$.

- Ta cần tìm thêm dạng biểu diễn cho $\vec{v}_3$ cũng là tổ hợp tuyến tính của $\vec{v}_1$ để:

  $$k\vec{v}_1 = \vec{v}_3$$

  Tuy nhiên $4 = \frac{4}{5} \times 5$ và $3 = \frac{3}{4} \times 4$. Do đó, $\vec{v}_3$ không thể là tổ hợp tuyến tính của $\vec{v}_1$ ($k\vec{v}_1 \neq \vec{v}_3$)

Vậy hạng của ma trận không thể bằng 1, lúc này hạng nhỏ nhất của ma trận A là 2. Chọn $k = \frac{1}{2}$ thì $\vec{v}_2 = \frac{1}{2}\vec{v}_1$. Lúc này, $\text{rank}(A) = 2$ với $\vec{v}_1, \vec{v}_3$ độc lập tuyến tính, a, d không âm và tùy ý, $b = \frac{3}{2}$ và $c = \frac{a}{2}$. Chọn một bộ thỏa là $(a, b, c, d) = (2, \frac{3}{2}, 1, 1)$

2.b

Do trong câu 2.a, ta cố gắng đưa $\vec{v}_2$ và $\vec{v}_3$ về dạng tổ hợp tuyến tính của vector $\vec{v}_1$. Để ma trận có hạng nhỏ nhất, ta thực hiện các phép biến đổi sơ cấp trên dòng sao cho trên ma trận A xuất hiện nhiều dòng toàn 0 nhất có thể. Thực chất điều này làm cho định thức của ma trận A bằng với 0. Vì, khi định thức của A bằng 0 thì ít nhất 1 trong các dòng của A là tổ hợp tuyến tính của 1 hay nhiều dòng khác. Do ma trận A trong bài không thể có hạng 1, ta đi tìm điều kiện tổng quát để a có hạng là 2. Tính định thức cho A theo qui tắc laplace cho hàng thứ nhất:

$$\begin{aligned}\det(A) &= \begin{vmatrix}4 & 3 & a\\ 2 & b & c\\ 5 & 4 & d\end{vmatrix}\\ &= 4\begin{vmatrix}b & c\\ 4 & d\end{vmatrix} - 3\begin{vmatrix}2 & c\\ 5 & d\end{vmatrix} + a\begin{vmatrix}2 & b\\ 5 & 4\end{vmatrix}\\ &= 4(bd-4c) - 3(2d-5c) + a(8-5b)\\ &= 4bd - 16c - 6d + 15c + a(8-5b)\\ &= 4bd - 6d - c + 8a - 5ab\end{aligned}$$

Để ma trận có hạng là 2 thì

$$\begin{aligned}\det(A) = 0 &\Leftrightarrow 4bd - 6d - c + 8a - 5ab = 0\\ &\Leftrightarrow d = \frac{5ab + c - 8a}{4b-6}, b \neq \frac{3}{2} \ \vee c = \frac{a}{2}, b = \frac{3}{2}\end{aligned}$$

Vậy để $\text{rank}(A) = 2$ thì bộ số phải có dạng:

- $(a, b, c, \frac{5ab+c-8a}{4b-6})$ với $a, b, c \in \mathbb{R}$ và $b \neq \frac{3}{2}$.
- $(a, \frac{3}{2}, \frac{a}{2}, d)$ với $a, d \in \mathbb{R}$ và $b = \frac{3}{2}$.

3

Gọi độ dài 3 cạnh hình hộp chữ nhật lần lượt là a, b, c với $a, b, c > 0$. Với mỗi hình hộp chữ nhật trong tập S gồm 3 cạnh a, b, c thì bán kính của hình cầu ngoại tiếp hình hộp chính là bán kính hình cầu nhỏ nhất có thể chứa được hình hộp. Ta có bán kính của hình cầu ngoại tiếp hình hộp cạnh a, b, c bất kì là

$$r = \frac{\sqrt{a^2+b^2+c^2}}{2}$$

Như vậy, ta cần tìm bán kính quả cầu ngoại tiếp hình hộp lớn nhất để có thể chứa được mọi quả cầu trong tập S thỏa

$$\begin{cases} abc = 23\\ ab + bc + ac = 27\end{cases}$$

Đặt $g(a, b, c) = abc - 23 = 0$ và $h(a, b, c) = ab + bc + ac - 27 = 0$.
Ta lại có, $(a^2 + b^2 + c^2) > 0$, do đó $r = \frac{\sqrt{a^2+b^2+c^2}}{2}$ lớn nhất thì $(a^2 + b^2 + c^2)$ cũng lớn nhất.

Ta xây dựng lại bài toán như sau:

$$\begin{aligned}&\max_{(a,b,c)} f(a,b,c) = a^2+b^2+c^2\\&\text{subject to:}\\&g(a,b,c) = abc - 23\\&h(a,b,c) = ab+bc+ac-27\end{aligned}$$

Gọi $D = \{a,b,c\} \in \mathbb{R}^3 \mid g(a,b,c) = h(a,b,c) = 0;\ a,b,c \geq 0\}$. Ta cần chứng minh tập D này là compact để chỉ ra f tồn tại giá trị lớn nhất. Ta dễ dàng thấy D là tập đóng vì giao của các tập đóng là tập đóng.

Theo định nghĩa, trong không gian $\mathbb{R}^n$, $D \subset \mathbb{R}^n$ bị chặn nếu tồn tại $r > 0$ sao cho $\|\vec{x}\| \leq r, \forall \vec{x} \in D$ (nghĩa là D nằm gọn trong một bán kính r nào đó).

Không mất tính tổng quát, ta giả sử $a \geq b \geq c \geq 0$.

Ta có $ab+bc+ac = 27 \Rightarrow ab+ac = a(b+c) \leq 27(*)$

Ta lại có $bc = \frac{23}{a} \Rightarrow b+c \geq 2\sqrt{bc} = 2\sqrt{\frac{23}{a}}$ (bất đẳng thức Cauchy). Nhân 2 vế cho a với $a \geq 0$ ta được: $a(b+c) \geq 2\sqrt{23a}(**)$

Từ $(*), (**)$ ta có: $\sqrt{23a} \leq a(b+c) \leq 27 \Rightarrow \sqrt{23a} \leq 27 \Rightarrow a \leq \frac{729}{92} \approx 7.924$

Như vậy $\frac{729}{92} \geq a \geq b \geq c \geq 0$, do đó $\forall a,b,c \in D$ thì $\|\vec{x}\| = \sqrt{a^2+b^2+c^2} \leq \sqrt{3 \times (\frac{729}{92})^2}$. Do đó D bị chặn. Như vậy ta chứng minh D đóng và D bị chặn. Vậy D compact.

Đặt $S(x) = span(\nabla g, \nabla h)$

Theo bổ đề 2, nếu $x^* = (a^*, b^*, c^*)$ là 1 cực trị của f trên D và $S(x^*) \neq 0$ thì $\nabla f(x^*) \in S(x^*)$. Vậy $\exists x^*, \lambda^*, \mu^*$ sao cho:

$$\nabla f(x^*) = \lambda^* \nabla g(x^*) + \mu^* \nabla h(x^*)$$

Xét:

$$\begin{cases}\nabla f_a = \lambda \nabla g_a + \mu \nabla h_a\\ \nabla f_b = \lambda \nabla g_b + \mu \nabla h_b\\ \nabla f_c = \lambda \nabla g_c + \mu \nabla h_c\end{cases} \tag{1}$$

$$\Leftrightarrow \begin{cases}2a = \lambda bc + \mu(b+c)\\ 2b = \lambda ac + \mu(a+c)\\ 2c = \lambda ab + \mu(a+b)\end{cases} \tag{2}$$

$$\Leftrightarrow \begin{cases}(a-b)(2+\lambda c+\mu) = 0\\ (b-c)(2+\lambda a+\mu) = 0\\ (a-c)(2+\lambda b+\mu) = 0\end{cases} \tag{3}$$

Nhìn vào biểu thức (3), vì a,b,c có vai trò như nhau, không mất tính tổng quát sẽ có 2 trường hợp xảy ra

$$a = b = c \text{ hoặc } a = b \neq c$$

Giả sử $a = b = c$, khi đó $abc = 23 \Leftrightarrow a = b = c = \sqrt[3]{23}$ thế vào $ab + bc + ac \neq 27$. Vậy $a = b = c$ là sai. Từ đó ta suy ra:

$$a = b = \frac{-\mu + 2}{\lambda} \neq c$$

Nhìn vào phương trình đầu tiên của biểu thức (2), ta chia 2 vế cho a, khi đó:

$$\begin{aligned} 2 &= \lambda c + \mu + \mu\frac{c}{a} \\ \Leftrightarrow c &= \frac{a(2-\mu)}{a\lambda + \mu} \end{aligned}$$

mà $a = \frac{-\mu+2}{\lambda}$ do đó

$$c = \frac{(\mu+2)(2-\mu)}{2\lambda}$$

Xét phương trình $g(a, b, c) = 0$, khi đó:

$$\begin{aligned} &abc = 23 \\ \Leftrightarrow &\frac{(\mu+2)^2}{\lambda^2}\frac{(\mu+2)(2-\mu)}{2\lambda} = 23 \\ \Leftrightarrow &\lambda = \frac{(u+2)\sqrt[3]{2-u}}{\sqrt[3]{46}} \end{aligned}$$

Thế λ vào a, b, c ta được

$$a = b = \frac{\sqrt[3]{46}}{\sqrt[3]{\mu-2}};\ c = \frac{\sqrt[3]{46(2-\mu)^2}}{2}$$

Xét phương trình $h(a, b, c) = 0$, khi đó:

$$\begin{aligned} &ab + bc + ac = 27 \\ \Leftrightarrow\ &\sqrt[3]{(\frac{46}{\mu-2})^2} + \sqrt[3]{46^2(2-\mu)} - 27 = 0 \\ \Leftrightarrow\ &(2-\mu)46^{\frac{2}{3}} + 27\sqrt[3]{(\mu-2)^2} - 46^{\frac{2}{3}} = 0 \\ \Leftrightarrow\ &\sqrt[3]{\mu-2} = 0.6074013106\ \vee\ \sqrt[3]{\mu-2} = -0.9189706907\ \vee\ \sqrt[3]{\mu-2} = -1.791523936 \\ \Leftrightarrow\ &\mu = 1.775907576\ \vee \mu = 2.918970691\ \vee \mu = \frac{31}{4} \end{aligned}$$

Vậy ta có được từng bộ nghiệm a, b, c tương ứng với μ:

$$a_1 = b_1 = 2;\ c_1 = \frac{23}{4}\ \ \mu_1 = \frac{31}{4}\ ; \vee \tag{4}$$

$$a_2 = b_2 = 3.68540644;\ c_2 = 1.693390251\ ; \mu_2 = 2.918970691 \vee \tag{5}$$

$$a_3 = b_3 = -5,898979484\ c_3 = 0.6609583059\ ; \mu_3 = 1.775907576 \tag{6}$$

Với a_3, b_3, c_3, vì a_3, b_3 mang giá trị âm nên ta không nhận bộ nghiệm này.
Với a_1, b_1, c_1 ta tính được $f(a_1, b_1, c_1) = \frac{657}{16} = 41.0625$
Với a_2, b_2, c_2 ta tính được $f(a_2, b_2, c_2) \approx 32.693$
Theo yêu cầu của bài toán, ta cần tìm bán kính lớn nhất của mặt cầu ngoại tiếp hình hộp chữ nhật. Mà $f(a_1, b_1, c_1) > f(a_2, b_2, c_2)$ do đó ta nhận bộ nghiệm $(a_1, b_1, c_1) = (2, 2, \frac{23}{4})$ với bán kính là

$$r = \frac{\sqrt{a^2 + b^2 + c^2}}{2} = \frac{\sqrt{2^2 + 2^2 + (\frac{23}{4})^2}}{2} = \frac{3\sqrt{73}}{8}$$

4

Tập dữ liệu $\theta = \{\vec{p_1} = (1, 3), \vec{p_2} = (0, 2), \vec{p_3} = (3, 4)\} \subset R^2$.

4.a

Ta sẽ chứng minh khoảng cách ngắn nhất từ $\vec{p} \in \mathbb{R}^n$ bất kì tới 1 hyperplane $W : \vec{w}^\top \vec{x} + b = 0$ với $W \subset \mathbb{R}^n$ và $\vec{w}, \vec{x} \in \mathbb{R}^n$, là khoảng cách vuông góc từ $\vec{p}$ đến hyperplane đó.
Xét $\vec{x} \in W$, khi đó ta phát biểu bài toán như sau:

$$\begin{aligned} &\min_{\vec{x}} f(\vec{x}) := \|\vec{p} - \vec{x}\|^2 \\ &\text{subject to: } \vec{w}^\top \vec{x} + b = 0 \end{aligned}$$

Gọi $g(\vec{x}) = \vec{w}^\top \vec{x} + b = 0$. Tập $D = \{\vec{x} \in \mathbb{R}^n \mid g(\vec{x}) = 0\}$. Ta sẽ chứng minh tập D là lồi.
Trước tiên, ta chứng minh tổng quát hyperplane trong không gian n chiều là $a_1x_1 + \cdots + a_nx_n + b = \vec{w}^\top \vec{x} + b = 0$ là lồi với $a_i, b, x_i \in \mathbb{R}$. Khi đó giả sử nếu

$$\vec{w}_1^\top \vec{x} + b = \vec{w}_2^\top \vec{x} + b = 0$$

thì với $t \in [0, 1]$ ta có:

$$t(\vec{w}_1^\top \vec{x} + b) + (1 - t)(\vec{w}_2^\top \vec{x} + b) = t(\vec{w}_1^\top \vec{x} - \vec{w}_2^\top \vec{x}) + t(-b + b) = t(-b - (-b)) + 0t = 0t + 0t = 0$$

Do đó, hyperplane trong không gian n chiều là các tập lồi. Vậy tập $D = \{\vec{x} \in \mathbb{R}^n \mid g(\vec{x}) = 0\}$ là lồi. Theo bổ đề 2, khi x^* là cực trị của f trên D và $\nabla g \neq 0$ thì tồn tại $(\vec{x}^*, \lambda^*)$ sao cho:

$$\nabla f(\vec{x}^*) = \lambda \nabla g(\vec{x}^*)$$

Xét:

$$\begin{aligned} &\nabla f(\vec{x}) = \lambda \nabla g(\vec{x}) \\ \Leftrightarrow &-2(\vec{p} - \vec{x}) = \lambda \vec{w} \\ \Leftrightarrow &\vec{x} = \vec{p} + \frac{\lambda}{2}\vec{w} \end{aligned}$$

Thế $\vec{x} = \vec{p} + \frac{\lambda}{2}\vec{w}$ vào $g(\vec{x})$ ta có:

$$\begin{aligned}
&\vec{w}^\top(\vec{p} + \frac{\lambda}{2}\vec{w}) + b = 0 \\
\Leftrightarrow &\lambda = -\frac{2(\vec{w}^\top\vec{p} + b)}{\|w\|^2}
\end{aligned}$$

Thế $\lambda = -\frac{2(\vec{w}^\top\vec{p}+b)}{\|w\|^2}$ vào $\vec{x} = \vec{p} + \frac{\lambda}{2}\vec{w}$ ta có:

$$\begin{aligned}
&\vec{x} = \vec{p} - \frac{\vec{w}^\top\vec{p} + b}{\|w\|^2}\vec{w} \\
\Leftrightarrow &\vec{p} - \vec{x} = \frac{\vec{w}^\top\vec{p} + b}{\|w\|^2}\vec{w}
\end{aligned}$$

Từ đây ta có thể kết luận:

$$\min_{\vec{x}} \|\vec{p} - \vec{x}\|^2 = \|\frac{\vec{w}^\top\vec{p} + b}{\|w\|^2}\vec{w}\|^2$$

là cực tiểu toàn cục của $\|\vec{p} - \vec{x}\|^2$.

Hơn nữa, theo kết quả từ bài tập về nhà 1, $\vec{x} = \operatorname{proj}_W \vec{p}$ là nghiệm của bài toán cực trị: Tìm $\vec{x} \in W$ sao cho $\|\vec{p} - \vec{x}\|$ đạt giá trị cực tiểu. Do đó, $\vec{x} = \operatorname{proj}_W \vec{p}$ là hình chiếu của $\vec{p}$ lên không gian con W, nên ta suy ra:

$$\frac{\vec{w}^\top\vec{p} + b}{\|w\|^2}\vec{w} \in W^\perp$$

với $W^\perp$ là phần bù trực giao của $W \subset \mathbb{R}^n$.

Vì thế, ta có bình phương khoảng cách vuông góc từ một điểm $\vec{p} \in \mathbb{R}^n$ trong đến hyperplane $\vec{w}^\top\vec{x} + b = 0$là:

$$\|\frac{\vec{w}^\top\vec{p} + b}{\|w\|^2}\vec{w}\|^2 = \frac{\|\vec{w}^\top\vec{p} + b\|^2}{\|w\|^2}$$

Bây giờ ta sẽ xây dựng hàm mất mát để tìm đường thẳng $d : 0 = ax - y + b$ sao cho tổng bình phương khoảng cách vuông góc từ các điểm trong θ đến d là nhỏ nhất.

$$L(w, b) = \sum_{i=1}^{N} \frac{\|\vec{w}^\top\vec{p}_i + b\|^2}{\|\vec{w}\|^2}$$

Ta lại có $\vec{w} = (a, -1)$ và tập dữ liệu $\theta = \{\vec{p}_1 = (1, 3), \vec{p}_2 = (0, 2), \vec{p}_3 = (3, 4)\} \subset \mathbb{R}^2$. Khi đó:

$$\begin{aligned}
L(a, -1, b) &= \frac{(a - 3 + b)^2 + (b - 2)^2 + (3a - 4 + b)^2}{a^2 + 1} \\
&= \frac{10a^2 + 3b^2 - 30a - 18b + 29 + 8ab}{a^2 + 1}
\end{aligned}$$

Đạo hàm riêng của L cho a:

$$\begin{aligned}\nabla L_a &= \frac{(20a - 30 + b)(a^2+1) - 2a(10a^2 + 3b^2 - 30a - 18b + 29 + 8a}{(a^2+1)^2} \\ &= \frac{30a^2 - 8a^2b - 38a - 6ab^2 + 36ab + 8b - 30}{(a^2+1)^2}\end{aligned}$$

Đạo hàm riêng của L cho b:

$$\nabla L_b = \frac{6b - 18 + 8a}{a^2+1}$$

Ta tìm điểm dừng bằng cách giải hệ:

$$\begin{aligned}&\begin{cases}\frac{\partial L}{\partial a} = 0\\ \frac{\partial L}{\partial b} = 0\end{cases}\\ \Leftrightarrow &\begin{cases}30a^2 - 8a^2b - 38a - 6ab^2 + 36ab + 8b - 30 = 0\\ 6b - 18 + 8a = 0\end{cases}\\ \Leftrightarrow &\begin{cases}6a^2 + \frac{16}{3}a - 6 = 0\\ b = 3 - \frac{4}{3}a\end{cases}\\ \Leftrightarrow &\begin{cases}a = \frac{-4+\sqrt{97}}{9} \vee a = \frac{-4-\sqrt{97}}{9}\\ b = 3 - \frac{4}{3}a\end{cases}\end{aligned}$$

Như vậy ta có hai bộ nghiệm $(a_1 = \frac{-4+\sqrt{97}}{9}, b_1 = \frac{97-4\sqrt{97}}{27})$ và $(a_2 = \frac{-4-\sqrt{97}}{9}, b_2 = \frac{97+4\sqrt{97}}{27})$
Thay a_1, b_1 vào L ta thấy bộ nghiệm này cho kết quả nhỏ hơn a_2, b_2.
Vậy bộ nghiệm $(a_1 = \frac{-4+\sqrt{97}}{9}, b_1 = \frac{97-4\sqrt{97}}{27})$ là nghiệm cho bài toán. Khi đó:

$$y = \frac{-4+\sqrt{97}}{9} \times x + \frac{97-4\sqrt{97}}{27}$$

4.b

Ta sẽ sử dụng định lí: Ma trận $B \in \mathbb{R}^{n\times n}$ đối xứng thì tồn tại một cơ sở trực giao $\mathbb{R}^n$ gồm các vector riêng của B.

1. Lập ma trận hiệp phương sai

Nếu $\{\vec{w}_1, \dots, \vec{w}_n\}$ là cơ sở trực giao của không gian vector V thì:

$$\text{proj}_{span(\vec{w}_1,\dots,\vec{w}_n)} \vec{v} = \text{proj}_{\vec{w}_1} \vec{v} + \dots + \text{proj}_{\vec{w}_n}$$

Cho $\theta = \{\vec{x}_1, \dots, \vec{x}_n\} \subset \mathbb{R}^n$ và $\vec{b}_1 \in \mathbb{R}^n$ thì độ khuếch tán theo phương $\vec{b}_1$ của θ là:

$$\text{Var}_{\vec{b}_1}(\theta) = \|\text{proj}_{\vec{b}_1} \vec{x}_1\|^2 + \dots + \text{proj}_{\vec{b}_1} \vec{x}_n\|^2$$

Phương $\vec{b}_1$ được gọi là thành phần chính thứ nhất của θ nếu $\underline{\mathrm{Var}_{\vec{b}_1}(\theta) \text{ đạt giá trị lớn nhất}}$
Giả sử θ là các cột của $\hat{X}$ và $\vec{b}_1$ là thành phần chính thứ nhất. Ta lập ma trận hiệp phương sai như sau:

$$A = \frac{1}{n}\hat{X}\hat{X}^\top$$

Giá trị $\frac{1}{n}$ trong A không ảnh hưởng tới kết quả bài toán. Khi đó, ma trận $A \in \mathbb{R}^{n\times n}$ đối xứng, tồn tại:

$$A\vec{b}_1 = \lambda_1\vec{b}_1$$

với λ_1 là trị riêng lớn nhất của A và $\vec{b}_1$ là vector riêng ứng với trị riêng λ_1 này.
Giả sử $\|\vec{b}_1\|^2 = 1$. Khi đó:

$$\begin{aligned}
\|\operatorname{proj}_{\vec{b}_1}\vec{x}_i\| &= (\operatorname{proj}_{\vec{b}_1}\vec{x}_i)\cdot(\operatorname{proj}_{\vec{b}_1}\vec{x}_i)\\
&= (\frac{\vec{b}_1\cdot\vec{x}_i}{\vec{b}_1\cdot\vec{b}_1}\cdot\vec{b}_1)\cdot(\frac{\vec{b}_1\cdot\vec{x}_i}{\vec{b}_1\cdot\vec{b}_1}\cdot\vec{b}_1)\\
&= [(\vec{b}_1\cdot\vec{x}_i)\cdot\vec{b}_1]^2\\
&= (\vec{b}_1\cdot\vec{x}_i)^2
\end{aligned}$$

Vậy:

$$\mathrm{Var}_{\vec{b}_1}(\theta) = \sum_{i=1}^{n}\operatorname{proj}_{\vec{b}_1}\vec{x}_i = \sum_{i=1}^{n}(\vec{b}_1\cdot\vec{x}_i)^2 = \left\|\begin{matrix}\vec{x}_1\cdot\vec{b}_1\\ \vdots\\ \vec{x}_n\cdot\vec{b}_1\end{matrix}\right\|^2 = \|\hat{X}^\top\vec{b}_1\|^2 = \vec{b}_1^\top\hat{X}\hat{X}^\top\vec{b}_1 = \vec{b}_1^\top A\vec{b}_1 = \left\langle\vec{b}_1, A\vec{b}_1\right\rangle$$

2. Tìm trị riêng và vector riêng của A

Bài toán:

$$\begin{aligned}
&\max_{\vec{b}_1}\mathrm{Var}_{\vec{b}_1}(\theta) = \vec{b}_1^\top A\vec{b}_1\\
&\text{subject to: } \|\vec{b}_1\|^2 = 1
\end{aligned}$$

Đặt $g(\vec{b}_1) = \|\vec{b}_1\|^2 - 1 = 0$. Tập $D = \{\vec{b}_1 \in \mathbb{R}^n \mid g(\vec{b}_1) = 0\}$. Khi đó nếu $\vec{b}_1^*$ là 1 cực trị của f trên D và $\nabla g = 2\vec{b}_1 \neq 0$ thì tồn tại $\lambda^*, \vec{b}_1^*$ sao cho:

$$\nabla f(\vec{b}_1^*) = \lambda^*\nabla g(\vec{b}_1^*)$$

Ta tìm điểm dừng bằng cách:

$$\begin{aligned}
&\nabla f(\vec{b}_1) = \lambda\nabla g(\vec{b}_1)\\
\Leftrightarrow\ &2A\vec{b}_1 = \lambda 2\vec{b}_1\\
\Leftrightarrow\ &A\vec{b}_1 = \lambda\vec{b}_1\\
\Rightarrow\ &\mathrm{Var}_{\vec{b}_1}(\theta) = \vec{b}_1^\top A\vec{b}_1 = \vec{b}_1^\top\lambda\vec{b}_1 = \lambda
\end{aligned}$$

Như vậy để maximize độ khuếch tán theo phương $\vec{b}_1$ của θ thì chúng ta chọn một vector cơ sở (vector riêng) ứng với trị riêng λ lớn nhất của ma trận hiệp phương sai A.
Để tìm ra thành phần chính thứ 2 thì $\vec{b}_2 \perp \vec{b}_1$ và maximize $\vec{b}_2^\top A\vec{b}_2$.
Tương tự cho thành phần chính thừ k + 1 thì $\vec{b}_{k+1} \perp \vec{b}_k$ và maximize $\vec{b}_{k+1}^\top A\vec{b}_{k+1}$.
Một cách làm khác thay vì tính nhân tử lagrange
Ta cũng có một cách tiếp cận khác để giải bài toán trên như sau:
Giả sử $B = \{\vec{w}_1, \dots, \vec{w}_n\}$ là 1 cơ sở trực chuẩn của $\mathbb{R}^n$ gồm các vector riêng của A ứng với trị riêng $\lambda_1 \geq \lambda_2 \geq \dots \lambda_n \geq 0$.
Khi đó: $\vec{b}_1 = \sum_{i=1}^n \alpha_i \vec{w}_i$ và $A\vec{w}_i = \lambda_i \vec{w}_i$.
Từ đó ta có $A\vec{b}_1 = A(\sum_{i=1}^n \alpha_i \vec{w}_i) = \sum_{i=1}^n \lambda_i \alpha_i \vec{w}_i$.
Thật vậy, $\mathrm{Var}_{\vec{b}_1}(\theta) = \left\langle \vec{b}_1, A\vec{b}_1 \right\rangle = \left\langle \sum_{i=1}^n \alpha_i \vec{w}_i, \sum_{j=1}^n \lambda_j \alpha_j \vec{w}_j \right\rangle$
Nhìn vào biểu thức trên, nếu $i \neq j$ thì $\vec{w}_i \cdot \vec{w}_j = 0$.
Suy ra $\mathrm{Var}_{\vec{b}_1}(\theta) = \sum_{i=1}^n \lambda_i \alpha_i \|\vec{w}_i\|^2 = \sum_{i=1}^n \lambda_i \alpha_i^2 \leq \lambda_1(\sum_{i=1}^n \alpha_i^2) = \lambda_1$. Do $\sum_{i=1}^n \alpha_i^2 = 1$.
Dấu '=' xảy ra khi và chỉ khi $A\vec{b}_1 = \lambda_1 \vec{b}_1$. Có nghĩa là $\vec{b}_1 = \vec{w}_1$, vector riêng ứng với trị riêng lớn nhất λ_1 của A
Như thế, khi tìm thành phần chính thứ 2 ta sẽ tìm sao cho $\mathrm{Var}_{\vec{b}_2}(\theta)$ lớn nhất và $\vec{b}_1 \perp \vec{b}_2$.
Cách làm trên cũng cho ra một kết quả với việc tính nhân tử lagrange.

Thực hiện PCA cho bộ dữ liệu θ

Dựa vào chứng minh trên, ta tìm thành phần chính thứ nhất và thứ hai cho θ.
Xếp $\vec{p}_i$ thành các các cột của ma trận $X \in \mathbb{R}^{2\times 3}$ như sau:

$$X = \begin{bmatrix} \vec{p}_1 & \vec{p}_2 & \vec{p}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 \\ 3 & 2 & 4 \end{bmatrix}$$

Khi đó $\mu = \frac{1}{3}\sum_{i=1}^3 \vec{p}_i = (\frac{4}{3}, 3)$. Đặt:

$$\hat{X} = \begin{bmatrix} \vec{p}_1 - \mu & \vec{p}_2 - \mu & \vec{p}_3 - \mu \end{bmatrix} = \begin{bmatrix} \frac{-1}{3} & \frac{-4}{3} & \frac{5}{3} \\ 0 & -1 & 1 \end{bmatrix}$$

Ta lập ma trận hiệp phương sai:

$$A = \frac{1}{3}\hat{X}\hat{X}^T = \frac{1}{3}\begin{bmatrix} -\frac{1}{3} & -\frac{4}{3} & \frac{5}{3} \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} -\frac{1}{3} & 0 \\ -\frac{4}{3} & -1 \\ \frac{5}{3} & 1 \end{bmatrix} = \begin{bmatrix} \frac{14}{9} & 1 \\ 1 & \frac{2}{3} \end{bmatrix}$$

Ta sẽ tìm các trị riêng của A do trị riêng của ma trận A là nghiệm của phương trình đặc trưng:

$$\begin{aligned}
&P_A(\lambda) := det(A - \lambda I_2) = 0\\
&\Leftrightarrow det(\begin{bmatrix} \frac{14}{9} - \lambda & 1 \\ 1 & \frac{2}{3} - \lambda \end{bmatrix}) = 0\\
&\Leftrightarrow \lambda^2 - \frac{20}{9}\lambda + \frac{11}{9} = 0\\
&\Leftrightarrow \lambda_1 = \frac{10 + \sqrt{97}}{9} \vee \lambda_2 = \frac{10 - \sqrt{97}}{9}
\end{aligned}$$

- Xét $\lambda_1 = \frac{10+\sqrt{97}}{9}$

Không gian riêng $E(\lambda_1)$ chính là nghiệm của hệ phương trình: $(A - \lambda_1 I_2)X = 0$

$$\begin{aligned}
\begin{bmatrix} \frac{14}{9} - \lambda_1 & 1 \\ 1 & \frac{2}{3} - \lambda_1 \end{bmatrix} &= \begin{bmatrix} \frac{4-\sqrt{97}}{9} & 1 \\ 1 & \frac{-4-\sqrt{97}}{9} \end{bmatrix}\\
&\to \begin{bmatrix} 4 - \sqrt{97} & 9 \\ 9 & -4 - \sqrt{97} \end{bmatrix}\\
&\to \begin{bmatrix} 4 - \sqrt{97} & 9 \\ 0 & 0 \end{bmatrix}
\end{aligned}$$

Vậy $(9t, (\sqrt{97} - 4)t)$ là nghiệm tổng quát của hệ $(A - \lambda_1 I_2)X = 0$.
Suy ra $E(\lambda_1)$ có $\dim E(\lambda_1) = 1$ với cơ sở:

$$B_1 = \{\vec{v}_1 = \begin{pmatrix} 9 \\ \sqrt{97} - 4 \end{pmatrix}\}$$

- Xét $\lambda_2 = \frac{10-\sqrt{97}}{9}$

Không gian riêng $E(\lambda_2)$ chính là nghiệm của hệ phương trình: $(A - \lambda_2 I_2)X = 0$

$$\begin{aligned}
\begin{bmatrix} \frac{14}{9} - \lambda_2 & 1 \\ 1 & \frac{2}{3} - \lambda_2 \end{bmatrix} &= \begin{bmatrix} \frac{4+\sqrt{97}}{9} & 1 \\ 1 & \frac{-4+\sqrt{97}}{9} \end{bmatrix}\\
&\to \begin{bmatrix} 4 + \sqrt{97} & 9 \\ 9 & -4 + \sqrt{97} \end{bmatrix}\\
&\to \begin{bmatrix} 4 + \sqrt{97} & 9 \\ 0 & 0 \end{bmatrix}
\end{aligned}$$

Vậy $(9t, (-4 - \sqrt{97})t)$ là nghiệm tổng quát của hệ $(A - \lambda_2 I_2)X = 0$.

Suy ra $E(\lambda_2)$ có $\dim E(\lambda_2) = 1$ với cơ sở:

$$B_2 = \{\vec{v}_2 = \begin{pmatrix} 9 \\ -4 - \sqrt{97} \end{pmatrix}\}$$

Ma trận $A \in \mathbb{R}^{n \times n}$ là ma trận hiệp phương sai đối xứng thì tồn tại một cơ sở trực giao $\mathbb{R}^n$ gồm các vector riêng của A là $\{\vec{v}_1, \vec{v}_2\}$. Ta có thể lấy bộ cơ sở trực chuẩn như sau:

$$B = \{\frac{\vec{v}_1}{\|\vec{v}_1\|}, \frac{\vec{v}_2}{\|\vec{v}_2\|}\} = \{(\frac{4+\sqrt{97}}{\sqrt{194+8\sqrt{97}}}, \frac{9}{\sqrt{194+8\sqrt{97}}}), (\frac{4-\sqrt{97}}{\sqrt{194-8\sqrt{97}}}, \frac{9}{\sqrt{194-8\sqrt{97}}})\}$$

Vậy các vector trong cơ sở trực chuẩn B là hai thành phần chính của θ

4.c

Nhìn vào câu 4.a, ta thấy vector pháp tuyến của của đường thẳng $d_1 : 0 = ax - y + b$ là $\vec{n}_1 = (-4 + \sqrt{97}, -9)$. Vậy vector chỉ phương của d_1 lúc này là $\vec{u}_1 = (9, -4 + \sqrt{97})$. Vector chỉ phương này cũng là vector cơ sở ứng với trị riêng λ_1 là $\vec{v}_1 = (9, -4 + \sqrt{97})$ ở câu 4.b.
Nhìn vào bộ nghiệm còn lại của câu 4.a, ta thấy vector pháp tuyến của của đường thẳng $d_2 : 0 = ax - y + b$ là $\vec{n}_2 = (-4 - \sqrt{97}, -9)$. Vậy vector chỉ phương của d_2 lúc này là $\vec{u}_2 = (9, -4 - \sqrt{97})$. Vector chỉ phương này cũng là vector cơ sở ứng với trị riêng λ_2 là $\vec{v}_2 = (9, -4 - \sqrt{97})$ ở câu 4.b.
Như vậy, ta có thể suy đoán được rằng bản chất của hai bài toán này là như nhau. Ta có nhận xét như sau:

- Ở câu 4.a. ta chỉ ra đường thẳng d khi đó ta cực tiểu bình phương khoảng cách từ bộ dữ liệu θ đến không gian W. Gọi $\tilde{p_i}$ là điểm dữ liệu sau khi chiếu $\vec{p_i}$ xuống không gian W, khi đó ta có thể hiểu bài 4.a theo một cách khác là: ta tìm cách chiếu $\vec{p_i}$ xuống không gian W mà sao cho các điểm dữ liệu giống với $\tilde{p_i}$ nhất có thể.

- Trong khi đó ở bài 4.b, thực hiện bài toán PCA là tìm các thành phần chính $\vec{b}_1, \vec{b}_2$ để maximize độ khuếch tán theo phương của các thành phần chính của θ, cũng có thể hiểu theo nghĩa là tìm một không gian vector span bởi các thành phần chính sao cho khi chiếu xuống không gian này thì các điểm dữ liệu giống dữ liệu gốc nhất có thể.

Đây chính là 2 cách nhìn của bài toán PCA, một là dưới "Projection Perspective"và hai là dưới "Maximum Variance Perspective"đã được đề cập trong cuốn "Math for Machine Learning".
Phần chứng minh tổng quát cho bài toán PCA đã được đề cập trong phần 4.b.