

PHÂN TÍCH MA TRẬN

THỰC HÀNH PHƯƠNG PHÁP SỐ CHO KHOA HỌC DỮ LIỆU

Buổi 2

Ngày 22 tháng 3 năm 2023

Liên hệ

GOOGLE CLASSROOM: **mrhvcvm**

TRỢ GIẢNG:

- ▶ Nguyễn Thị Kiều Trang: ntktrang@hcmus.edu.vn
- ▶ Lý Như Bình: lnbinh@hcmus.edu.vn

LƯU Ý:

- ▶ Email đăng nhập google classroom thể hiện đầy đủ họ và tên, tránh sử dụng email có biệt danh.
- ▶ Tiêu đề mail (bắt buộc):
[2023-HK2-THPPSKHDL] [Tiêu đề thư]
VD: [2023-HK2-THPPSKHDL] HỎI BÀI
Vui lòng giới thiệu họ tên, MSSV và tên ca học khi gửi email.

Một vài điều về lớp

Điểm thực hành: Chiếm 40% tổng điểm:

- ▶ Điểm danh: 1 điểm (Mỗi buổi)
- ▶ Bài tập: 3 điểm (Nộp bài tập thực hành mỗi tuần)

Cách thức nộp bài:

- ▶ Nộp trên google classroom
- ▶ Nộp file .txt
- ▶ Tên file: Y_MSSV_Hoten_baix.txt,
 - ▶ $Y = LB$ nếu bạn học phòng C204.
 - ▶ $Y = LT$ nếu bạn học phòng C203.
 - ▶ $x \in \{1, 2, 3, 4, \dots\}$

Phân tích suy biến

Phân tích suy biến là phương pháp phân tích ma trận A kích thước $m \times n$ thành

$$A = U\Sigma V^T, \quad (1)$$

trong đó

- ▶ U là ma trận trực giao có kích thước $m \times m$.
- ▶ Σ là ma trận đường chéo có kích thước $m \times n$.
- ▶ V là ma trận trực giao có kích thước $n \times n$.

Phép giảm chiều của phân tích suy biến

Compact SVD:

Giả sử trên đường chéo chính có r phần tử khác không

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r > 0. \quad (2)$$

Phân tích compact SVD

$$A = U_r \Sigma_r V_r^T, \quad (3)$$

trong đó

- ▶ U_r và V_r là ma trận được tạo từ r cột đầu tiên của U và V .
- ▶ Σ_r là ma trận được tạo từ r hàng và r cột đầu tiên của Σ .

Phép giảm chiều của phân tích suy biến

Truncated SVD:

Trong ma trận Σ , các giá trị kỳ dị σ nằm trên đường chéo chính thường không âm và giảm dần.

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0. \quad (4)$$

Thông thường, chỉ một số ít các giá trị kỳ dị σ_i mang giá trị lớn, các giá trị còn lại thường nhỏ và gần 0. Khi đó, với phương pháp truncated SVD, ta có thể xấp xỉ ma trận A bằng tổng của $k < r$ ma trận có hạng 1

$$A \approx A_k = U_k \Sigma_k V_k^T. \quad (5)$$

Phép giảm chiều của phân tích suy biến

Ngoài ra, ta cũng có thể biểu diễn dưới dạng tổng của tích vô hướng các vector cột và dòng của U, V^T như sau:

$$A \approx A_k = \sum_{i=1}^k \sigma_i u_i v_i. \quad (6)$$

Ta có:

$$\|A - A_k\|_F^2 = \sum_{i=k+1}^r \sigma_i^2 \quad (7)$$

$$\text{với } \|X\|_F = \sqrt{\max_i^m \max_j^n |x_{ij}|^2} = \sqrt{\text{tr}(X^T X)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(X)} \quad (8)$$

Phép giảm chiều của phân tích suy biến

Ta biết rằng k giá trị kỳ dị σ được giữ lại có liên quan đến lượng thông tin chúng ta có thể giữ lại sau khi cắt giảm ma trận.

Tỉ lệ lượng thông tin có thể giữ lại sau khi cắt giảm $r - k$ giá trị kỳ dị của ma trận A là:

$$1 - \frac{\|A - A_k\|_F^2}{\|A\|_F^2} = \frac{\sum_{i=1}^k \sigma_i^2}{\sum_{j=1}^r \sigma_j^2} \quad (9)$$

Bài 1:

Thời gian: 30 phút.

Cho ma trận A sau đây:

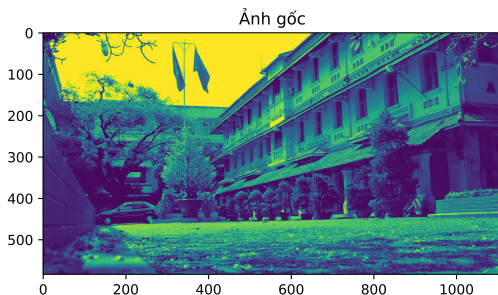
$$A = \begin{bmatrix} 11.08 & 6.82 & 1.76 & -6.82 \\ 2.50 & -1.01 & -2.60 & 1.19 \\ -4.88 & -5.07 & -3.21 & 5.20 \\ -0.49 & 1.52 & 2.07 & -1.66 \\ -14.04 & -12.40 & -6.66 & 12.65 \\ 0.27 & -8.51 & -10.19 & 9.15 \\ 9.53 & -9.84 & -17.00 & 11.00 \\ -12.01 & 3.64 & 11.10 & -4.48 \end{bmatrix} \quad (10)$$

Bài 1 tiếp theo

- a. Sử dụng hàm **svd** của package **scipy** trong Python để viết một chương trình cho phép tính dạng phân tích suy biến của ma trận A .
- b. Tìm ma trận compact SVD của ma trận A .
- c. Tìm ma trận truncated SVD của ma trận A sao cho có thể giữ lại ít nhất 85% lượng thông tin trong A .
- d. Tìm ma trận truncated SVD của ma trận A khi giữ lại 3 giá trị kỳ dị khác 0 của ma trận A . Sau đó, hãy tính lượng thông tin được giữ lại lúc này.

Bài 2: nén ảnh

Thời gian: 40 phút. Cho hình dưới đây:



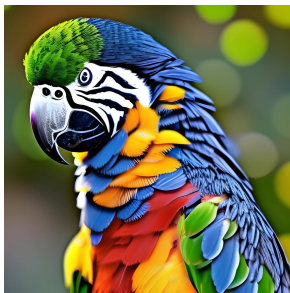
Hình 1: Hình gốc

Bài 2: tiếp theo

- a. Lưu hình ảnh trên về máy và dùng hàm `Image.open()` trong thư viện PIL của Python để mở.
- b. Chuyển hình ảnh vừa mở phía trên thành mảng bằng hàm `np.array()`.
- c. Chuyển hình ảnh trên sang grayscale.
- d. Xem hình ảnh trên là một ma trận. Hãy viết hàm con cho phép truyền vào số giá trị kỳ dị được giữ lại và xuất ra hình ảnh đã nén theo phương pháp truncated SVD của hình ban đầu.
- e. Hãy xuất ảnh được nén tương ứng nếu ta lần lượt giữ 25, 50, 75 và 150 số giá trị kỳ dị của hình ảnh ban đầu. Tính tỷ lệ lượng thông tin được giữ lại sau khi đã nén tương ứng của từng trường hợp.

Bài 3:

Thời gian: 40 phút. Cho hình dưới đây:



Hình 2: Hình gốc

Bài 3: tiếp theo

- a. Lưu hình ảnh trên về máy và dùng hàm `Image.open()` trong thư viện PIL của Python để mở.
- b. Chuyển hình ảnh vừa mở phía trên thành mảng bằng hàm `np.array()`.
- c. Chuyển hình ảnh trên sang grayscale.
- d. Xem hình ảnh trên là một ma trận. Hãy viết hàm con cho phép truyền vào số giá trị kỳ dị được giữ lại và xuất ra hình ảnh đã nén theo phương pháp truncated SVD của hình ban đầu.
- e. Hãy xuất ảnh được nén tương ứng nếu ta lần lượt giữ 10, 25, 100, và 150 số giá trị kỳ dị của hình ảnh ban đầu. Tính tỷ lệ lượng thông tin được giữ lại sau khi đã nén tương ứng của từng trường hợp.

Bài toán bình phương tối thiểu

$$\min_x \|Ax - b\|_2^2 \quad (11)$$

Nếu $A^T A$ khả nghịch, nghiệm của bài toán $x = (A^T A)^{-1} A^T b$.

Nếu $A^T A$ không khả nghịch, nghiệm của bài toán

$$x^+ = A^+ b = V \Sigma^+ U^T b$$

Bài 4:

Thời gian: 30 phút Giả sử ta có số liệu thống kê các thông số RAM, bộ nhớ, PIN và giá tiền của 24 chiếc điện thoại như trong bảng

	NAME	RAM	MEMORY	PIN	PRICE
0	Samsung Galaxy A01 Core	1	16	3000	1,850,000
1	Samsung Galaxy A11	3	32	4000	2,650,000
2	Samsung Galaxy A02s	4	64	5000	3,350,000
3	Samsung Galaxy J7 Prime	3	32	3300	3,790,000
4	Samsung Galaxy A21s	3	32	5000	4,250,000
5	Samsung Galaxy A22 4G	6	128	5000	4,700,000
6	Samsung Galaxy A30s	4	64	4000	4,150,000
7	Samsung Galaxy A31	6	128	5000	5,150,000
8	Samsung Galaxy A52	8	128	4500	6,600,000
9	Samsung Galaxy A72	8	256	5000	10,100,000
10	Samsung Galaxy Note 10 Lite	8	128	4500	10,500,000
11	Samsung Galaxy S10+	8	128	4100	12,400,000

	NAME	RAM	MEMORY	PIN	PRICE
12	Samsung Galaxy S21 FE 5G	8	256	4500	13,650,000
13	Samsung Galaxy S21 FE	8	128	4500	12,790,000
14	Samsung Galaxy S20+	8	128	4500	15,500,000
15	Samsung Galaxy S20 Ultra	12	128	5000	16,000,000
16	Samsung Galaxy Note 20 Ultra	12	256	4500	18,990,000
17	Samsung Galaxy Z Flip3 5G	8	256	3300	19,350,000
18	Samsung Galaxy Z Flip	8	256	3300	20,990,000
19	Samsung Galaxy S21 Plus 5G	8	256	4800	23,000,000
20	Samsung Galaxy Z Fold2	12	256	4500	23,000,000
21	Samsung Galaxy S21 Ultra 5G	12	256	5000	29,800,000
22	Samsung Galaxy S22 Ultra	12	512	5000	29,990,000
23	Samsung Galaxy Z Fold3	12	512	4400	33,990,000

Hình 3: Bảng số liệu thống kê số RAM, bộ nhớ, PIN và giá tiền của 24 chiếc điện thoại.

Bài 3 tiếp theo

Viết chương trình để tìm giá trị dự đoán giá tiền của một chiếc điện thoại mới dựa vào các thông số RAM, bộ nhớ và PIN lần lượt là 4, 64, 4000.