ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN KHOA HỌC MÁY TÍNH

SINGULAR VALUE DECOMPOSITION - SVD

KHAI THÁC DỮ LIỆU VÀ ỨNG DỤNG

Lê Nhựt Nam

Ngày 18 tháng 10 năm 2024

MỤC LỤC

MA TRẬN SVD

Giới thiệu

Các ma trận đối xứng luôn luôn chéo hóa được (trực giao).
Nghĩa là, với *bất kỳ ma trận đối xứng* $\mathbf{A} \in \mathbb{R}^{n \times n}$, *tồn tại một ma trận trực giao* $\mathbf{Q} = [q_1 ... q_n]$ *và một ma trận đường chéo* $\mathbf{\Lambda} = diag(\lambda_1, ..., \lambda_n)$, *đều là ma trận số thực vuông, sao cho*:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top. \tag{1}$$

Ta đã chỉ ra rằng λ_i là các trị riêng của $\mathbf{A}$ và q_i là các vector riêng tương ứng (trực giao với nhau và có chuẩn bằng 1).
Do đó, phân tích này được gọi là phân tích trị riêng của $\mathbf{A}$, hay còn gọi là phân tích phổ của $\mathbf{A}$.

Câu hỏi 1
Thế với các ma trận hình chữ nhật tổng quát thì sao?

Sự tồn tại của SVD cho ma trận tổng quát

Định lý 1

Với bất kỳ ma trận $\mathbf{X} \in \mathbb{R}^{n \times d}$, tồn tại hai ma trận trực giao $\mathbf{U} \in \mathbb{R}^{n \times n}$ và $\mathbf{V} \in \mathbb{R}^{d \times d}$ và một ma trận đường chéo $\Sigma \in \mathbb{R}^{n \times d}$ (có cùng kích thước với ma trận $\mathbf{X}$) sao cho:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top. \tag{2}$$

Nhận xét 1

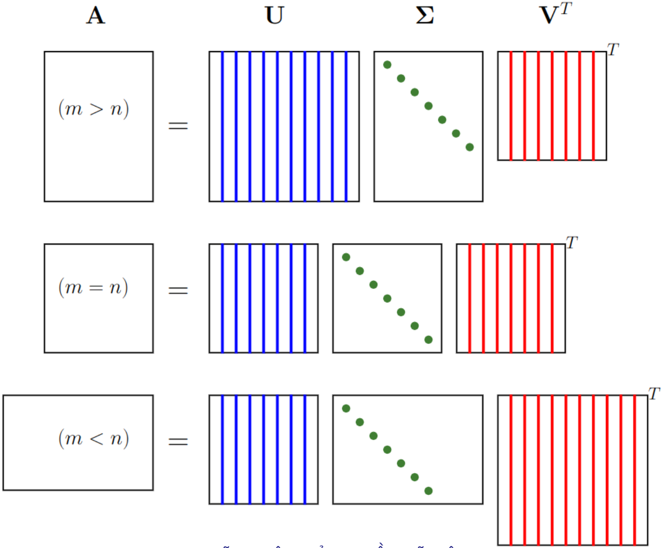
Đây gọi là *Singular Value Decomposition (SVD)* of $\mathbf{X}$:

(i) Đường chéo của $\mathbf{\Sigma}$ được gọi là các giá trị kỳ dị (singular values)

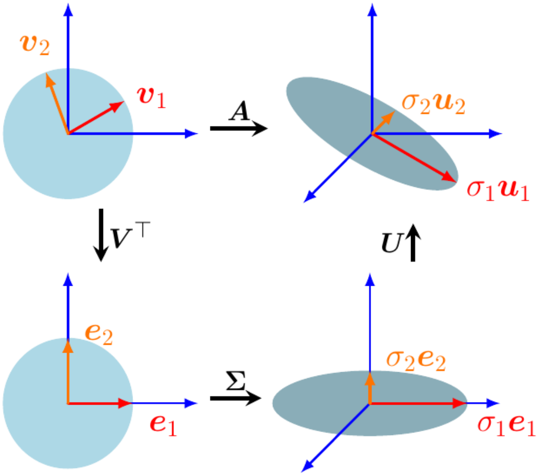
(ii) Các cột của $\mathbf{U}$ được gọi là các vector kỳ dị trái (left singular vectors).

(iii) Các cột của $\mathbf{V}$ được gọi là các vector kỳ dị phải (right singular vectors).

Minh họa SVD hoàn chỉnh



Minh họa SVD hoàn chỉnh



Ví dụ

Ví dụ 1

Dễ dàng ta có:

$$\underbrace{\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}}_{\mathbf{U}} \cdot \underbrace{\begin{pmatrix} \sqrt{3} & \\ & 1 \\ & \end{pmatrix}}_{\mathbf{\Sigma}} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{\top}}_{\mathbf{V}^{\top}}$$

.

Ví dụ

- Singular values:

$$\sigma_1 = \sqrt{3}, \quad \sigma_2 = 1$$

- Left singular vectors:

$$\mathbf{u}_1 = \begin{pmatrix} \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}$$

- Right singular vectors:

$$\mathbf{v}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

Liên hệ đến phân rã phổ của các ma trận đối xứng

Từ SVD của **X** ta được:

$$\mathbf{X}\mathbf{X}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \cdot \mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top = \mathbf{U}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)\mathbf{U}^\top \tag{3}$$

$$\mathbf{X}^\top\mathbf{X} = \mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top \cdot \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = \mathbf{V}(\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma})\mathbf{V}^\top \tag{4}$$

Điều này cho thấy:

(i) $\mathbf{U}$ là ma trận vector riêng của $\mathbf{X}\mathbf{X}^\top$;

(ii) $\mathbf{V}$ là ma trận vector riêng của $\mathbf{X}^\top\mathbf{X}$;

(iii) Các trị riêng của $\mathbf{X}\mathbf{X}^\top$ và $\mathbf{X}^\top\mathbf{X}$ (phải giống nhau) bằng với <u>bình phương</u> của các giá trị kỳ dị của $\mathbf{X}$.

Ví dụ

Ví dụ 2

Với ma trận **A** trong Ví dụ trước.

$$\mathbf{A}\mathbf{A}^\top = \underbrace{\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}}_{\mathbf{U}} \cdot \underbrace{\begin{pmatrix} 3 & & \\ & 1 & \\ & & 0 \end{pmatrix}}_{\mathbf{\Sigma}^\top} \cdot \underbrace{\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}^\top}_{\mathbf{U}^\top}$$

$$\mathbf{A}^\top\mathbf{A} = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\mathbf{V}} \cdot \underbrace{\begin{pmatrix} 3 & 1 \end{pmatrix}}_{\mathbf{\Sigma}^\top\mathbf{\Sigma}} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^\top}_{\mathbf{V}^\top}$$

Chứng minh định lý SVD

Với bất kỳ ma trận $\mathbf{X} \in \mathbb{R}^{n \times d}$, SVD có thể được coi là việc giải một phương trình ma trận cho ba ma trận chưa biết (mỗi ma trận có những ràng buộc nhất định):

$$\mathbf{X} = \underbrace{\mathbf{U}}_{\text{trực giao}} \cdot \underbrace{\mathbf{\Sigma}}_{\text{đường chéo}} \cdot \underbrace{\mathbf{V}^\top}_{\text{trực giao}}. \tag{5}$$

Giả sử tồn tại những ma trận như vậy. Ta đã chứng minh được

$$\mathbf{X}^\top \mathbf{X} = \mathbf{V} \mathbf{\Sigma}^\top \mathbf{\Sigma} \mathbf{V}^\top.$$

Điều này cho chúng ta biết cách tìm $\mathbf{V}$ và $\mathbf{\Sigma}$ (bao gồm các vector riêng và căn bậc hai của các trị riêng của $\mathbf{X}^\top \mathbf{X}$, tương ứng).

Chứng minh định lý SVD

Sau khi tìm được $\mathbf{V}$ và $\mathbf{\Sigma}$, viết lại biểu thức thành:

$$\mathbf{XV} = \mathbf{U\Sigma}.$$

hay dưới dạng cột

$$\mathbf{X}\left[\mathbf{v}_1 \dots \mathbf{v}_r \; \mathbf{v}_{r+1} \dots \mathbf{v}_d\right] = \left[\mathbf{u}_1 \dots \mathbf{u}_r \; \mathbf{u}_{r+1} \dots \mathbf{u}_n\right] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & & \end{bmatrix}$$

Bằng cách so sánh các cột, ta được:

$$\mathbf{X}_i = \begin{cases} \sigma_i \mathbf{u}_i, & 1 \leq i \leq r \text{ (\#singular value khác 0)} \\ 0, & r < i \leq d \end{cases}$$

Từ đó tìm được ma trận $\mathbf{U}$: $\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{X}\mathbf{v}_i$ for $1 \leq i \leq r$.

Chứng minh định lý SVD

Với $\mathbf{C} = \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{d \times d}$. Ma trận $\mathbf{C}$ là ma trận vuông, đối xứng và nửa xác định dương. Do đó, theo định lý phổ, $\mathbf{C} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$ với $\mathbf{V} \in \mathbb{R}^{d \times d}$ là ma trận trực giao và $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$ là ma trận đường chéo với $\lambda_1 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_d$ (trong đó $r = \operatorname{rank}(\mathbf{X}) \leq d$). Giả sử $\sigma_i = \sqrt{\lambda_i}$ và từ đó tạo thành ma trận $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times d}$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \operatorname{diag}(\sigma_1, \ldots, \sigma_r) & \mathbf{O}_{r \times (d-r)} \\ \mathbf{O}_{(n-r) \times r} & \mathbf{O}_{(n-r) \times (d-r)} \end{bmatrix}$$

Thêm vào đó:

$$\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{X} \mathbf{v}_i \in \mathbb{R}^n, \qquad \text{với mỗi } 1 \leq i \leq r$$

Chứng minh định lý SVD

Có $\mathbf{u}_1, \ldots, \mathbf{u}_r$ là các vector trực chuẩn. Để thấy điều này,

$$\begin{aligned}
\mathbf{u}_i^\top \mathbf{u}_j &= \left(\frac{1}{\sigma_i}\mathbf{X}\mathbf{v}_i\right)^\top \left(\frac{1}{\sigma_j}\mathbf{X}\mathbf{v}_j\right) = \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^\top \underbrace{\mathbf{X}^\top\mathbf{X}}_{=\mathbf{C}}\mathbf{v}_j \\
&= \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^\top(\lambda_j\mathbf{v}_j) = \frac{\sigma_j}{\sigma_i}\mathbf{v}_i^\top\mathbf{v}_j \quad (\lambda_j = \sigma_j^2) \\
&= \begin{cases} 1, & \text{nếu } i = j \\ 0, & \text{nếu } i \neq j \end{cases}
\end{aligned}$$

Chọn $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n \in \mathbb{R}^n$ (dựa trên đổi cơ sở) sao cho:

$$\mathbf{U} = [\mathbf{u}_1 \ldots \mathbf{u}_r \mathbf{u}_{r+1} \ldots \mathbf{u}_n] \in \mathbb{R}^{n\times n}$$

là một ma trận trực giao.

Chứng minh định lý SVD

Việc còn lại là chứng minh $\mathbf{XV} = \mathbf{U\Sigma}$, i.e.,

$$\mathbf{X}\left[\mathbf{v}_1 \ldots \mathbf{v}_r \ \mathbf{v}_{r+1} \ldots \mathbf{v}_d\right] = \left[\mathbf{u}_1 \ldots \mathbf{u}_r \ \mathbf{u}_{r+1} \ldots \mathbf{u}_n\right] \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \\ & & & \end{bmatrix}$$

Xét 2 trường hợp:

- $1 \leq i \leq r : \mathbf{X}\mathbf{v}_i = \sigma_i \mathbf{u}_i$.
- $i > r : \mathbf{X}\mathbf{v}_i = \mathbf{0}$, điều này xảy ra do $\mathbf{X}^\top \mathbf{X}\mathbf{v}_i = \mathbf{C}\mathbf{v}_i = 0\mathbf{v}_i = \mathbf{0}$.

Từ đó, ta được $\mathbf{X} = \mathbf{U\Sigma V}^\top$.

Các biến thể của SVD

(a) **SVD hoàn chỉnh (Full SVD):** $\mathbf{X}_{n\times d} = \mathbf{U}_{n\times n}\boldsymbol{\Sigma}_{n\times d}\mathbf{V}_{d\times d}^{\top}$

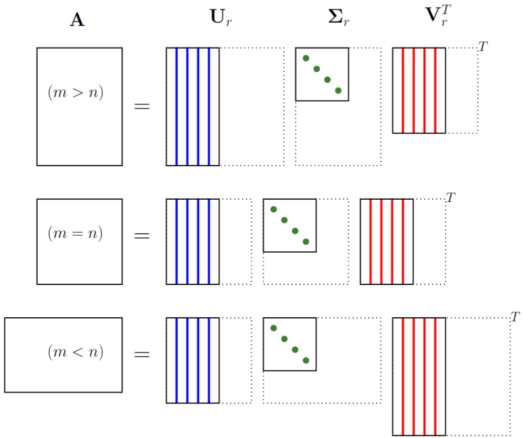
(b) **SVD tinh gọn (Compact SVD):** Giả sử $\text{rank}(\mathbf{X}) = r$. Định nghĩa

$$\mathbf{U}_r = [\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{n\times r}$$
$$\mathbf{V}_r = [\mathbf{v}_1, \dots, \mathbf{v}_r] \in \mathbb{R}^{d\times r}$$
$$\boldsymbol{\Sigma}_r = \text{diag}\left(\sigma_1, \dots, \sigma_r\right) \in \mathbb{R}^{r\times r}$$

Từ đó, ta có:

$$\mathbf{X} = \mathbf{U}_r\boldsymbol{\Sigma}_r\mathbf{V}_r^{\top}$$

Minh họa Compact SVD



Các biến thể của SVD

(c) **Phân tích Rank-1**:

$$\mathbf{X} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_r^\top \end{bmatrix} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

Điều này có thể hiểu là $\mathbf{X}$ là tổng trọng số của các ma trận rank-1, như đối với ma trận vuông, đối xứng $\mathbf{A} \in \mathbb{R}^{n \times n}$:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top = \sum_{i=1}^{n} \lambda_i \mathbf{q}_i \mathbf{q}_i^\top$$

Tóm lại, $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^\top$ với cả $\mathbf{U}$ và $\mathbf{V}$ đều có các cột trực chuẩn và $\mathbf{\Sigma}$ là ma trận đường chéo.

Các biến thể của SVD

Nhận xét 2

Đối với bất kỳ phiên bản nào của SVD, hình thức không phải là duy nhất (điều này chủ yếu do các lựa chọn khác nhau của cơ sở trực giao cho mỗi không gian riêng).

Các biến thể của SVD

Nhận xét 3

Đối với bất kỳ ma trận $\mathbf{X} \in \mathbb{R}^{n \times d}$ và số nguyên $1 \leq \mathbf{K} \leq r$, chúng ta định nghĩa SVD bị cắt của $\mathbf{X}$ với $\mathbf{K}$ hạng như sau

$$\mathbf{X} \approx \sum_{i=1}^{\mathbf{K}} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = \mathbf{X}_\mathbf{K}$$

trong đó các trị riêng được giả định sắp xếp từ lớn đến nhỏ (vì vậy $\sigma_1, \ldots, \sigma_\mathbf{K}$ đại diện cho $\mathbf{K}$ trị riêng lớn nhất).

Các biến thể của SVD

Lưu ý 1

Lưu ý rằng $\mathbf{X_K}$ có hạng bằng $\mathbf{K}$ và không hoàn toàn bằng $\mathbf{X}$ (do đó có thể được coi là một xấp xỉ của $\mathbf{X}$).

Ý nghĩa hình học của SVD đầy đủ

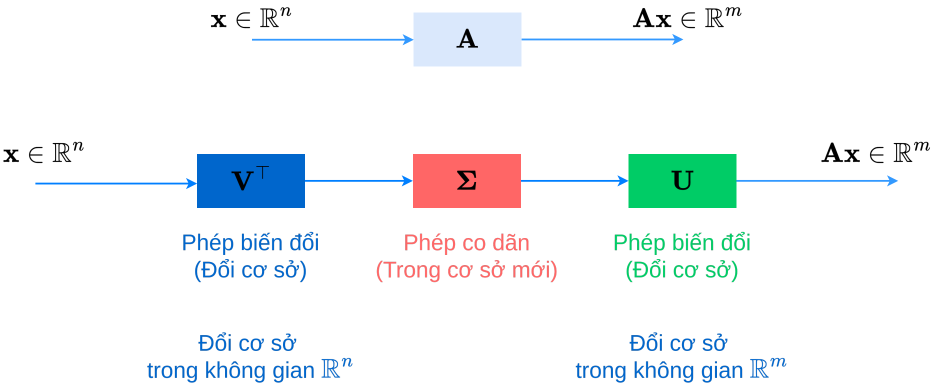
Cho bất kỳ ma trận $\mathbf{A} \in \mathbb{R}^{m \times n}$ xác định một phép biến đổi tuyến tính:

$$f : \mathbb{R}^n \mapsto \mathbb{R}^m, \quad \text{với} \quad f(\mathbf{x}) = \mathbf{A}\mathbf{x}$$

SVD của $\mathbf{A}$ chỉ ra rằng phép biến đổi tuyến tính f có thể được phân tách thành một chuỗi ba thành phần:

$$\underbrace{\mathbf{A}\mathbf{x}}_{\text{biến đổi hoàn chỉnh}} = \underbrace{\mathbf{U}}_{\text{phép xoay}} \cdot \underbrace{\boldsymbol{\Sigma}}_{\text{phép co dãn}} \cdot \underbrace{\mathbf{V}^\top\mathbf{x}}_{\text{phép xoay}}$$

Ý nghĩa hình học của SVD đầy đủ



Co dãn dọc theo cơ sở mới bởi các singular values

(i) Nếu $m < n$ - Bỏ $n - m$ cơ sở cuối (tác dụng của các cột 0 trong $\mathbf{\Sigma}$)

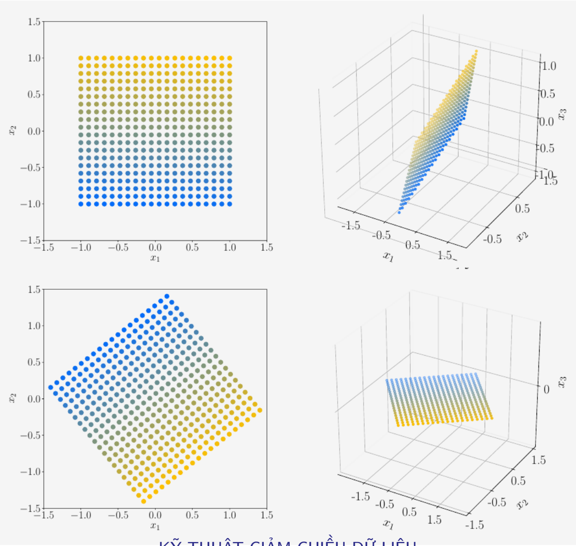
(ii) Nếu $m > n$ - Thêm $m - n$ cơ sở (tác dụng của các dòng 0 trong $\mathbf{\Sigma}$)

Ý nghĩa hình học của SVD đầy đủ

Ví dụ 3

$$\mathbf{A} = \begin{pmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \mathbf{U\Sigma V}^\top = \begin{pmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{pmatrix} \begin{pmatrix} 1.62 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{pmatrix}$$

Ý nghĩa hình học của SVD đầy đủ



Ý nghĩa hình học của Compact SVD

Cho trước bất kỳ ma trận $\mathbf{A} \in \mathbb{R}^{m \times n}$ có hạng r, phân rã SVD compact của nó là:

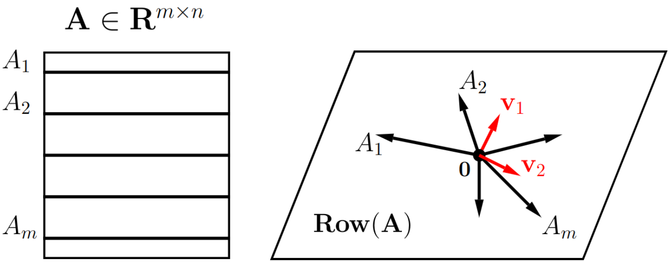
$$\mathbf{A} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^\top$$

Dạng SVD compact có thể được viết dưới dạng:

$$\begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} = \mathbf{A} = \underbrace{(\mathbf{U}_r \mathbf{\Sigma}_r)}_{\text{ma trận hệ số}} \cdot \underbrace{\mathbf{V}_r^\top}_{\text{cơ sở}} = \begin{bmatrix} * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^\top \\ \mathbf{v}_2^\top \\ \vdots \\ \mathbf{v}_r^\top \end{bmatrix}.$$

Ý nghĩa hình học của Compact SVD

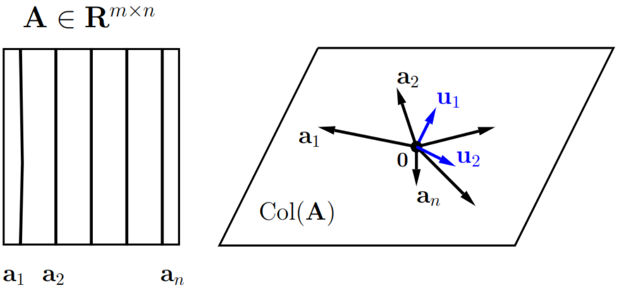
Các dòng của ma trận $\mathbf{V}_r^\top$ (hay các cột của ma trận $\mathbf{V}_r$) cho ta một cơ sở trực chuẩn (orthonormal) với không gian dòng của ma trận $\mathbf{A}$.



Ý nghĩa hình học của Compact SVD

Tương tự, các cột của ma trận $\mathbf{U}_r$ cho ta một cơ sở trực chuẩn (orthonormal) với không gian cột của ma trận $\mathbf{A}$

$$\mathbf{A} = \underbrace{\mathbf{U}_r}_{\text{cơ sở}} \cdot \underbrace{\left(\boldsymbol{\Sigma}_r \mathbf{V}_r^T\right)}_{\text{các hệ số}} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r \end{bmatrix} \cdot \left(\boldsymbol{\Sigma}_r \mathbf{V}_r^T\right).$$



Phương pháp lũy thừa cho tính toán số học SVD

Giả sử $\mathbf{A} \in \mathbb{R}^{m \times n}$ là một ma trận mà SVD của nó sẽ được tính toán: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$. Xét $\mathbf{C} = \mathbf{A}^\top\mathbf{A} \in \mathbb{R}^{n \times n}$. Khi đó,

$$
\begin{aligned}
\mathbf{C} &= \mathbf{V}\left(\mathbf{\Sigma}^\top\mathbf{\Sigma}\right)\mathbf{V}^\top = \sum \sigma_i^2 \mathbf{v}_i\mathbf{v}_i^\top \\
\mathbf{C}^2 &= \mathbf{V}\left(\mathbf{\Sigma}^\top\mathbf{\Sigma}\right)^2\mathbf{V}^\top = \sum \sigma_i^4 \mathbf{v}_i\mathbf{v}_i^\top \\
&\vdots \\
\mathbf{C}^k &= \mathbf{V}\left(\mathbf{\Sigma}^\top\mathbf{\Sigma}\right)^k\mathbf{V}^\top = \sum \sigma_i^{2k} \mathbf{v}_i\mathbf{v}_i^\top
\end{aligned}
$$

Nếu $\sigma_1 > \sigma_2$, thì hạng tử đầu tiên chiếm ưu thế, vì vậy $\mathbf{C}^k \to \sigma_1^{2k}\mathbf{v}_1\mathbf{v}_1^\top$ khi $k \to \infty$. Điều này có nghĩa là một ước lượng gần đúng cho $\mathbf{v}_1$ có thể được tính bằng cách đơn giản là lấy cột đầu tiên của $\mathbf{C}^k$ và chuẩn hóa nó thành một vector đơn vị.

Phương pháp lũy thừa cho tính toán số học SVD

Nhận xét 4

Phương pháp trước đây rất tốn kém do phần lũy thừa ma trận.

Một cách tiếp cận tốt hơn. Thay vì tính $\mathbf{C}^k$, chúng ta chọn một vector ngẫu nhiên $\mathbf{x} \in \mathbb{R}^n$ và tính $\mathbf{C}^k\mathbf{x}$ thông qua một chuỗi các phép nhân ma trận-vector (các phép toán này rất hiệu quả, đặc biệt khi một chiều của $\mathbf{A}$ nhỏ hoặc $\mathbf{A}$ thưa):

$$\mathbf{C}^k\mathbf{x} = \mathbf{A}^\top\mathbf{A}\cdots\mathbf{A}^\top\mathbf{A}\mathbf{x}$$

Với $\mathbf{x} = \sum c_i\mathbf{v}_i$ (vì $\mathbf{v}_1, \ldots, \mathbf{v}_n$ tạo thành một cơ sở trực chuẩn cho $\mathbb{R}^n$). Khi đó:

$$\mathbf{C}^k\mathbf{x} \approx \left(\sigma_1^{2k}\mathbf{v}_1\mathbf{v}_1^\top\right)\left(\sum c_i\mathbf{v}_i\right) = \sigma_1^{2k}c_1\mathbf{v}_1$$

Chuẩn hóa vector $\mathbf{C}^k\mathbf{x}$ cho một giá trị lớn của k cho ra $\mathbf{v}_1$, vector đơn vị riêng thứ nhất của $\mathbf{A}$.

VÍ DỤ CÁCH TÍNH SVD

Bài toán 1

Cho ma trận

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix}$$

Tìm SVD của $\mathbf{A}$, tức là $\mathbf{U\Sigma V}^\top$.

Bài toán 1

Bước 1: Tính chuyển vị của ma trận $\mathbf{A}$ và $\mathbf{A}^\top\mathbf{A}$

$$\mathbf{A}^\top = \begin{pmatrix} 2 & 3 \\ 3 & 2 \\ -2 & 2 \end{pmatrix}$$

$$\mathbf{A}^\top\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 3 & 2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{pmatrix} = \begin{pmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{pmatrix}$$

Bài toán 1

Bước 2: Xác định các trị riêng của $\mathbf{A}^\top\mathbf{A}$ và sắp xếp chúng theo thứ tự giảm dần. Các singular values của $\mathbf{A}$ là căn bậc hai của các giá trị này.

Đa thức đặc trưng $\det(\mathbf{A}^\top\mathbf{A} - \lambda\mathbf{I}) = \lambda^2 - 34\lambda + 225 = (\lambda - 25)(\lambda - 9)$.

Ta thấy hai trị riêng là $\lambda_1 = 25$ và $\lambda_2 = 9$. Nên hai singular values của $\mathbf{A}$ là $\sigma_1 = \sqrt{\lambda_1} = \sqrt{25} = 5$ và $\sigma_2 = \sqrt{\lambda_2} = \sqrt{9} = 3$.

Bài toán 1

Bước 3: Xây dựng ma trận đường chéo $\mathbf{\Sigma}$ bằng cách đặt các singular values giảm dần dọc theo đường chéo chính. Sau đó tính toán nghịch đảo của nó, tức là $\mathbf{\Sigma}^{-1}$.

$$\mathbf{\Sigma} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}$$

Bài toán 1

Bước 4: Sử dụng các trị riêng đã sắp thứ tự ở Bước 2 và tính toán các vector riêng của $\mathbf{A}^\top\mathbf{A}$. Thay thế các vector riêng này thành các cột của $\mathbf{V}$ và tính toán chuyển vị của nó. Với $\lambda_1 = 25$, ta có:

$$\mathbf{A}^\top\mathbf{A} - \lambda_1\mathbf{I}_3 = \begin{pmatrix} 13-25 & 12 & 2 \\ 12 & 13-25 & -2 \\ 2 & -2 & 8-25 \end{pmatrix} = \begin{pmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{pmatrix}$$

$$(\mathbf{A}^\top\mathbf{A} - \lambda_1\mathbf{I}_3)v_1 = 0$$

$$\begin{pmatrix} -12 & 12 & 2 \\ 12 & -12 & -2 \\ 2 & -2 & -17 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$$

Bài toán 1

$$\mathbf{A}^\top\mathbf{A} - \lambda_2\mathbf{I}_3 = \begin{pmatrix} 13-9 & 12 & 2 \\ 12 & 13-9 & -2 \\ 2 & -2 & 8-9 \end{pmatrix} = \begin{pmatrix} 4 & 12 & 2 \\ 12 & 4 & -2 \\ 2 & -2 & -1 \end{pmatrix}$$

$$(\mathbf{A}^\top\mathbf{A} - \lambda_1\mathbf{I}_3)v_2 = 0$$

$$\begin{pmatrix} 4 & 12 & 2 \\ 12 & 4 & -2 \\ 2 & -2 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 0 & -\frac{1}{4} \\ 0 & 0 & -\frac{1}{4} \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow v_2 = \begin{pmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{pmatrix}$$

Bài toán 1

Do ta đang tính toán các vector singular trái (tức các cột của V), nên ta xét trị riêng $\lambda_3 = 0$. Ta có thể tính toán thông qua $\mathbf{A}^\top \mathbf{A}$ hoặc tìm một vector đơn vị vuông gốc với v_1 và v_2. Để mà vuông góc với $v_1 = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$, ta cần $-a = b$. Thì, ta có:

$v_1^\top v_3 = 0 \Leftrightarrow 2a/\sqrt{18} + 4c\sqrt{18} = 0 \Leftrightarrow -a = 2c$. Do đó:

$$v_3 = \begin{pmatrix} a \\ -a \\ -a/2 \end{pmatrix}$$

Và để nó là vector đơn vị, $a = 2/3$. Thế nên,

$$v_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ -1/3 \end{pmatrix}$$

Bài toán 1

Bước 5: Tính toán **U**

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = \mathbf{U}\begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \\ 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \end{pmatrix}^\top$$

Ta có thể tính toán **U** bằng cách $\sigma_i u_i = \mathbf{A}v_i \Leftrightarrow u_i = \frac{1}{\sigma_i}\mathbf{A}v_i$.

$$\mathbf{U} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

Bài toán 1

Như vậy,

$$\mathbf{A} = \mathbf{U\Sigma V}^\top = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \\ 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \end{pmatrix}^\top$$

Bài toán 2

Bài toán

Tính toán SVD đầy đủ cho ma trận sau

$$\mathbf{A} = \begin{pmatrix} 4 & 0 \\ 3 & -5 \end{pmatrix}$$

Bài toán 3

Bài toán

Tính toán SVD đầy đủ cho ma trận sau

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$

ỨNG DỤNG CỦA SVD

Phân rã ma trận

Cho một ma trận $\mathbf{A} \in \mathbb{R}^{m \times n}$, nó có thể được phân rã bằng cách sử dụng SVD

$$\mathbf{A} = \mathbf{U\Sigma V}^\top = \sum_{i=1}^{\min(m,n)} u_i \sigma_i v_i^\top$$

Phân rã ma trận - Cắt cụt

Nếu hạng của một ma trận là $r < \min(m, n)$ thì ta có thể cắt cụt tại r

$$\mathbf{A} = \sum_{i=1}^{r} u_i \sigma_i v_i^\top$$

Phân rã ma trận - xấp xỉ hạng k

Bằng cách sử dụng SVD, ta định nghĩa xấp xỉ hạng k như sau:

$$\mathbf{A} \approx \sum_{i=1}^{k} u_i \sigma_i v_i^\top$$

High-dimensional data

Dimensionality of data can be high, and even higher than number of samples. Consider dimensionality $D = 1\mathrm{M}$ (one million) and number of samples $N = 100$. All analysis still applies, but it would be wasteful to compute eigenvectors for the $1\mathrm{M} \times 1\mathrm{M}$ matrix, as its rank will anyway be at most N (thus 100). Let us define $\mathbf{X}$ to be a matrix formed by stacking all the data vectors (after having subtracted the mean from them): $\mathbf{X} = [\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, \ldots, \mathbf{x}_N - \bar{\mathbf{x}}]$
Thus,

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{T}} = \frac{1}{N} \mathbf{X}\mathbf{X}^{\mathrm{T}}$$

The characteristic equation is then

$$\frac{1}{N} \mathbf{X}\mathbf{X}^{\mathrm{T}} \mathbf{u} = \lambda \mathbf{u}$$

Left-multiplying both sides by $\mathbf{X}^{\mathrm{T}}$ gives

$$\frac{1}{N} \mathbf{X}^{\mathrm{T}} \mathbf{X} \overbrace{\left(\mathbf{X}^{\mathrm{T}} \mathbf{u}\right)}^{\mathbf{w}} = \lambda \overbrace{\left(\mathbf{X}^{\mathrm{T}} \mathbf{u}\right)}^{\mathbf{w}}$$

High-dimensional data

Thus, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, which is only 100×100, has exactly the same set of eigenvalues:

$$\frac{1}{N}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} = \lambda\mathbf{w}$$

Left-multiplying now by $\mathbf{X}$, we get

$$\frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{w}) = \lambda(\mathbf{X}\mathbf{w})$$

Conclusion: If $D \gg N$, form the matrix $\mathbf{T} = \frac{1}{N}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and compute its eigenvalues λ 's and eigenvectors $\mathbf{w}$. Compute the eigenvectors of $\mathbf{S} = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{T}}$ as

$$\mathbf{v} = \frac{\mathbf{X}\mathbf{w}}{\|\mathbf{X}\mathbf{w}\|}$$

Ví dụ: bộ dữ liệu Yale