Lecture 7

Announcements

- HW 2 due
- HW 3 posted

Readings

Strang I.9

Today

- SVD
- Matrix norms

Singular Value Decomposition

- For a real $m \times n$ matrix A, we can decompose $A = U \Sigma V^T$ such that
  - U is an $m \times m$ orthogonal matrix of left singular vectors
  - V is an $n \times n$ orthogonal matrix of right singular vectors
  - Σ is an $m \times n$ matrix containing singular values ...

The "usual" shape of SVD for data analysis

$$\underbrace{\begin{bmatrix} & & \\ & A & \\ & & \\ & & \end{bmatrix}}_{m \times n} = \underbrace{\begin{bmatrix} & & & \\ & U & & \\ & & & \\ & & & \end{bmatrix}}_{m \times m} \underbrace{\begin{bmatrix} & & \\ & \Sigma & \\ & & \\ & & \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} & V^T & \end{bmatrix}}_{n \times n}$$

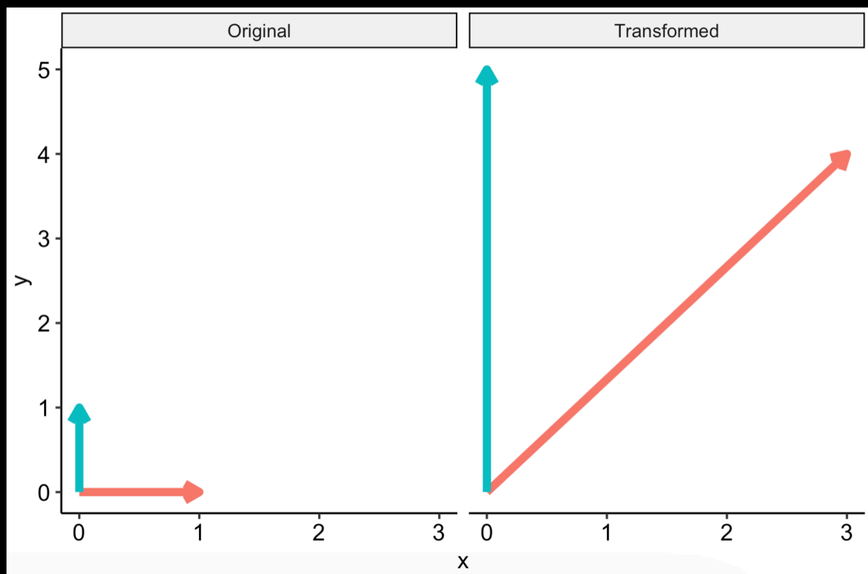
classical statistics: $m > n$

$$\begin{bmatrix} & & & \\ & A & & \\ & & & \end{bmatrix} = \underbrace{\begin{bmatrix} U \end{bmatrix}}_{m \times m} \underbrace{\begin{bmatrix} & & & \\ & \Sigma & & \\ & & & \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} & & & \\ & V^T & & \\ & & & \end{bmatrix}}_{n \times n}$$

"high dimensional" $n > m$

Geometric interpretation of SVD

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 3\sqrt{5} & 0 \\ 0 & \sqrt{5} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{V^T}$$



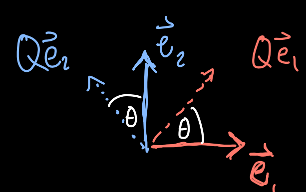
When $n = 2$, orthogonal matrices are either rotations or reflections.

* $\cos^2\theta + \sin^2\theta = 1$

Rotation

$$Q_{\text{rotate}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

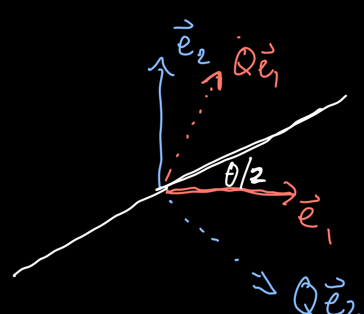
(rotation by angle θ)



Reflection

$$Q_{\text{reflect}} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

(reflection across the $\frac{\theta}{2}$ line)



Consider any 2×2 orthogonal matrix

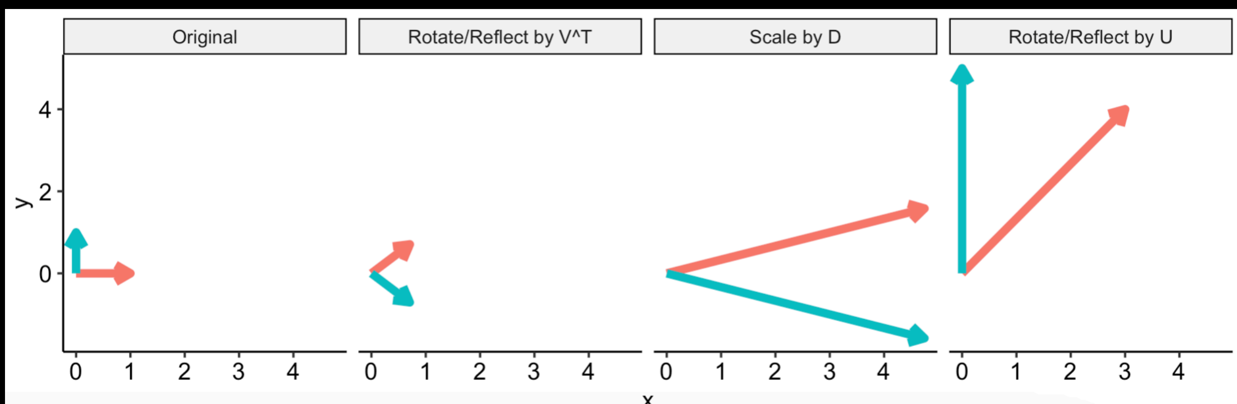
$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} \qquad -1 \le a \le 1$$

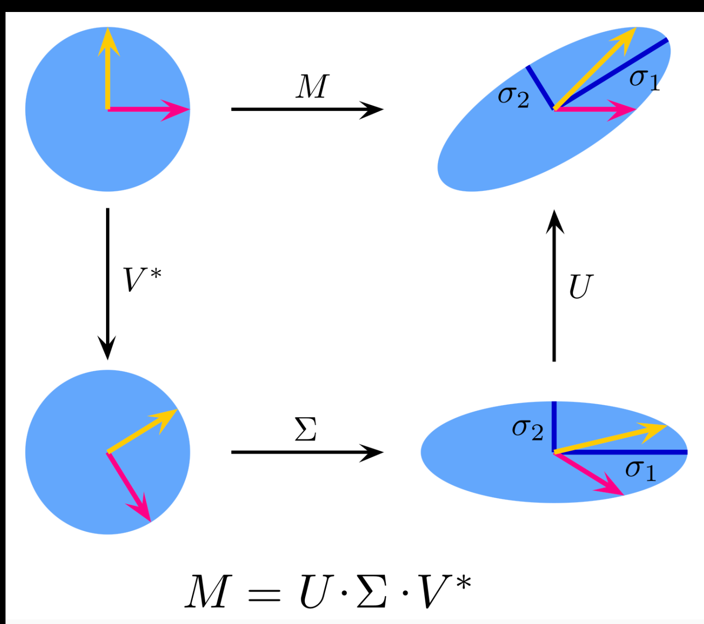
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{OR} \quad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

For SVD, we have

$$A = U \quad \Sigma \quad V^T$$

(rotation/reflection) x (scaling) x (rotation/reflection)





Properties of SVD

- $A = U\Sigma V^T$

- $A\vec{v}_i = \sigma_i \vec{u}_i$ for $i = 1, \dots, r$ $\quad r = \text{rank}(A)$

  $A\vec{v}_{r+1} = \dots A\vec{v}_n = \vec{0}$

- By convention, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$
- In general if A is rank r,

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \dots + \sigma_r \vec{u}_r \vec{v}_r^T$$

(sum of r rank-1 matrices)

- We can write A in a "reduced" SVD form

$$A = U_r \Sigma_r V_r^T$$

$$\underset{m\times n}{\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}} = \underset{m \times r}{\begin{bmatrix} & & \\ & U_r & \\ & & \end{bmatrix}} \underset{r\times r}{\overset{\Sigma_r}{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}}} \underset{r\times n}{\begin{bmatrix} & & \\ & V_r^T & \\ & & \end{bmatrix}}$$

note $V_r \in \mathbb{R}^{n \times r}$

and $V_r^T V_r = I_r$

$V_r V_r^T \neq I_n$

- $A_k = \sigma_1 \vec{u}_1 \vec{v}_1^T + \dots + \sigma_k \vec{u}_k \vec{v}_k^T$ is the "best" rank-k approximation to A. (Eckart-Young)

"Proof" of existence of SVD

$A = U\Sigma V^T$

What is A^TA?

$(U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma U^T U \Sigma V^T$

$= V\Sigma^2 V^T$

What is AA^T?

$AA^T = U\Sigma^2 U^T$

Key fact: AA^T and A^TA are both symmetric. (exercise: convince yourself)

Therefore $A^TA = V \Lambda V^T$ are both diagonalizable

$AA^T = U \Lambda U^T$

- V contains orthonormal eigenvectors of A^TA
- U contains orthonormal eigenvectors of AA^T
- σ_1^2 to σ_r^2 are the nonzero eigenvalues of A^TA and AA^T

(verify nonzero eigenvalues of AB and BA are the same)

Construction

① Choose orthonormal eigenvectors of A^TA $\vec{v}_1, \dots, \vec{v}_r$

② Choose $\sigma_k = \sqrt{\lambda_k}$

③ Set $\vec{u}_k = \dfrac{A\vec{v}_k}{\sigma_k}$

④ Choose $\vec{u}_k$ to be orthogonal for $k = r+1, \dots, m$ to complete U

Choose $\vec{v}_k$ to be orthogonal for $k = r+1, \dots, n$ to complete V

ex

$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ $A^TA = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$ $AA^T = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$

① eigenvalues of A^TA $\longrightarrow$ solve $\det(A^TA - \lambda I) = 0$

$\lambda_1 = 45 = \sigma_1^2$

$\lambda_2 = 5 = \sigma_2^2$

② solve for eigenvectors

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 45 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \vec{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$\begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 5 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Longrightarrow \qquad \vec{v}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

③ Solve for U

$$A\vec{v}_1 = \sigma_1 \vec{u}_1 \Rightarrow \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \sqrt{45} \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix}$$

$$= \underset{\sigma_1}{\sqrt{45}} \underset{\vec{u}_1}{\begin{bmatrix} 1/\sqrt{10} \\ 3/\sqrt{10} \end{bmatrix}}$$

$$A\vec{v}_2 = \sigma_2 \vec{u}_2 = \underset{\sigma_2}{\sqrt{5}} \underset{\vec{u}_2}{\begin{bmatrix} -3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}}$$

Remarks

① If S is symmetric, its SVD is $Q \Lambda Q^T = U \Sigma V^T$ and all singular values are positive.

② If $A = Q$ is orthogonal, what is its SVD?

$A^T A = I \Rightarrow$ all singular values are 1 and $\Sigma = I$

③ Without loss of generality, we can choose singular values to be non negative (by changing signs of singular vectors)

Eckart–Young

If we compute $A = U\Sigma V^T = \sigma_1 \vec{u}_1 \vec{v}_1^T + \dots + \sigma_r \vec{u}_r \vec{v}_r^T$,

then $A_k = \sigma_1 \vec{u}_1 \vec{v}_1^T + \dots + \sigma_k \vec{u}_k \vec{v}_k^T$ $\quad (k \le r)$

is the "best" rank-k approximation to A.

Matrix norms

Review: vector norms

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

(p-norm)

For any vector norm, we can define an operator norm.

def Let $\|\cdot\|$ be any vector norm.

The corresponding operator norm is

$$\|A\| = \sup_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|}{\|\vec{x}\|}$$

$$\|Q\|_2 = \sup_{\vec{x} \neq \vec{0}} \frac{\|Q\vec{x}\|_2}{\|\vec{x}\|_2} = 1$$

note

- $\|A\| \geq 0$ for all A
- $\|A\| = 0$ if and only if $A = 0$
- $\|\alpha A\| = |\alpha| \, \|A\|$ for all real numbers α
- $\|A + B\| \le \|A\| + \|B\|$ (triangle inequality)
- quantify distance between A and B using $\|A - B\|$

Eckart–Young

$\|A - A_k\|$ is minimized when A_k is defined as above.