vv214: Singular Value Decomposition. Low Rank Approximations.

Dr.Olga Danilkina

UM-SJTU Joint Institute

July 23, 2020

1. SVD
2. Frobenius norm
3. Low Rank Approximations
4. Pseudoinverses

SVD: Motivation

Let $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ be linear, $L\bar{x} = A\bar{x}$.

$$(A^T A)^T = A^T (A^T)^T = A^T A \Rightarrow A^T A \text{ is symmetric}$$

$$\text{Spectral Theorem} \Rightarrow \exists \text{ an orthonormal basis } \bar{v}_1,\ \bar{v}_2 \text{ for } A^T A$$

$$(A\bar{v}_1, A\bar{v}_2) = (A\bar{v}_1)^T A\bar{v}_2 = \bar{v}_1^T A^T A\bar{v}_2 = \bar{v}_1^T \lambda_2 \bar{v}_2 = \lambda_2(\bar{v}_1,\ \bar{v}_2) = 0$$

$$\Rightarrow A\bar{v}_1 \perp A\bar{v}_2$$

SVD: Example

Let $L\colon \mathbb{R}^2 \to \mathbb{R}^2$, $L\bar{x} = A\bar{x}$, $A = \begin{pmatrix} 6 & 2 \\ -7 & 6 \end{pmatrix}$

$$A^T A = \begin{pmatrix} 6 & -7 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 6 & 2 \\ -7 & 6 \end{pmatrix} = \begin{pmatrix} 85 & -30 \\ -30 & 40 \end{pmatrix}$$

Eigenvalues: $(85 - \lambda)(40 - \lambda) - 900 = 0 \Rightarrow \lambda_1 = 100,\ \lambda_2 = 25$

Orthonormal eigenbasis: $\bar{v}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix} \perp \bar{v}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$$A\bar{v}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 10 \\ -20 \end{pmatrix} \perp A\bar{v}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 10 \\ 5 \end{pmatrix}$$

SVD: Example

Consider the unit circle

$$\bar{x} = \bar{v}_1 \cos t + \bar{v}_2 \sin t$$

What is its image under the linear map L?

$$L\bar{x} = A\bar{v}_1 \cos t + A\bar{v}_2 \sin t,\ A\bar{v}_1 \perp A\bar{v}_2$$

$$\Rightarrow L\bar{x} \text{ is an ellipse with semi-axes } ||A\bar{v}_1||,\ ||A\bar{v}_2||$$

$$||A\bar{v}_1||^2 = (A\bar{v}_1, A\bar{v}_1) = \lambda_1(\bar{v}_1,\ \bar{v}_1) = \lambda_1 \Rightarrow ||A\bar{v}_1|| = \sqrt{100} = 10$$

$$||A\bar{v}_2|| = \sqrt{25} = 5$$

The eigenvalues of $A^T A$ define the ellipse as the image of the unit circle.

Singular Values

Definition: The singular values of a matrix $A_{n\times m}$ are the square roots of the eigenvalues of the symmetric matrix $(A^T A)_{m\times m}$ listed with their algebraic multiplicities:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m \geq 0$$

Theorem: Let $L\colon \mathbb{R}^2 \to \mathbb{R}^2$, $L\bar{x} = A\bar{x}$ be invertible. The image of the unit circle under the map L is an ellipse E. Singular values of A are the length of semi-axes of E.

Example

$$L\colon \mathbb{R}^3 \to \mathbb{R}^2,\ L\bar{x} = A\bar{x} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \bar{x}$$

$$A^T A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

$$(1-\lambda)^2(2-\lambda) - 2 = 0 \Rightarrow \lambda_1 = 3,\ \lambda_2 = 1,\ \lambda_3 = 0$$

Singular Values are $\quad \sigma_1 = \sqrt{3} > \sigma_2 = 1 > \sigma_3 = 0$

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

Example

$$A\bar{v}_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 3 \\ 3 \end{pmatrix}, A\bar{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}, A\bar{v}_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$||A\bar{v}_1|| = \sqrt{3} = \sigma_1,\ ||A\bar{v}_2|| = 1 = \sigma_2,\ ||A\bar{v}_3|| = 0 = \sigma_3$$

The unit sphere in $\mathbb{R}^3$ is defined by

$$\bar{x} = c_1\bar{v}_1 + c_2\bar{v}_2 + c_3\bar{v}_3, \quad c_1^2 + c_2^2 + c_3^2 = 1$$

The image of the unit sphere is

$$L\bar{x} = c_1 L\bar{v}_1 + c_2 L\bar{v}_2 = c_1\lambda_1\bar{v}_1 + c_2\lambda_2\bar{v}_2$$

$$c_1^2 + c_2^2 \leq 1$$

an ellipse

Singular Value Decomposition

Lemma: If $rank\, A_{n \times m} = r$, then its singular values

$$\sigma_1, \ldots, \sigma_r \neq 0 \quad \text{and} \quad \sigma_{r+1}, \ldots, \sigma_m = 0$$

Theorem (SVD): Any matrix $A_{n \times m}$ can be represented in the form

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

U is an orthogonal $n \times n$ matrix, V is an orthogonal $m \times m$ matrix

Σ is a matrix whose first r diagonal entries are

nonzero singular values of A, $r = rank\, A$, and all other entries vanish

Singular Value Decomposition

Singular Value Decomposition: Remarks

Remark 1:

$$A\bar{v}_i = \sigma_i \bar{u}_i,\ i = 1, \dots, r \qquad A\bar{v}_i = \bar{0},\ i = r+1, \dots, m$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$Im\,A = span(\bar{u}_1, \dots, \bar{u}_r) \qquad Ker\,A = span(\bar{v}_{r+1}, \dots, \bar{v}_m)$$

Remark 2: $A = U\Sigma V^T \Rightarrow A^T = V\Sigma^T \underbrace{U^T}_{U^{-1}} \Rightarrow A^T U = V\Sigma^T$

$$A^T\bar{u}_i = \sigma_i \bar{v}_i,\ i = 1, \dots, r \qquad A^T\bar{u}_i = \bar{0},\ i = r+1, \dots, m$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$Im\,A^T = span(\bar{v}_1, \dots, \bar{v}_r) \qquad Ker\,A^T = span(\bar{u}_{r+1}, \dots, \bar{u}_m)$$

Singular Value Decomposition: Example 1

$$A = \begin{pmatrix} 6 & 2 \\ -7 & 6 \end{pmatrix}$$

$$\bar{v}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \; \bar{v}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Rightarrow V = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

$$\bar{u}_1 = \frac{1}{\sigma_1} A \bar{v}_1 = \frac{1}{10\sqrt{5}} \begin{pmatrix} 10 \\ -20 \end{pmatrix}, \; \bar{u}_2 = \frac{1}{\sigma_2} A \bar{v}_2 = \frac{1}{5\sqrt{5}} \begin{pmatrix} 10 \\ 5 \end{pmatrix}$$

$$\Rightarrow U = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}$$

$$A = \begin{pmatrix} 6 & 2 \\ -7 & 6 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$$

Singular Value Decomposition: Example 2

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

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

$$\bar{u}_1 = \frac{1}{\sigma_1} A\bar{v}_1 = \frac{1}{\sqrt{3}\sqrt{6}} \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \ \bar{u}_2 = \frac{1}{\sigma_2} A\bar{v}_2 = \frac{1}{1 \cdot \sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\Rightarrow U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \Sigma = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \\ 1/\sqrt{2} & 0 & -1\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix}$$

MATLAB Commands

- texttts = svd(A) returns the singular values of matrix A in descending order.
- texttt[U,S,V] = svd(A) performs a singular value decomposition of matrix A, such that $A = U * S * V'$.
- texttts = svds(A,k) returns the k largest singular values.

The Frobenius Norm

The Frobenius norm of a matrix $A_{n\times m}$ which is defined by

$$|A||_F = \sqrt{\sum_{i,j=1}^{n,m} a_{ij}^2} = \sqrt{trace(AA^T)} = \sqrt{trace(A^TA)},$$

can be also represented by

$$||A||_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}, \text{ where } \sigma_i \text{ are singular values of } A$$

Condition Number of a Matrix

- Singular values $\sigma_1 \geq \sigma_2 \geq \sigma_n$ of a matrix $A_{n \times n}$ show how much distortion can occur under the linear transformation defined by A.
- Let $\exists A^{-1}$ and $S_2 = \{\bar{x} : ||\bar{x}||_2 = 1\}$ be a unit sphere in $\mathbb{R}^n$.
- $A = U \Sigma V^T$, $D = diag(\sigma_1, \ldots, \sigma_n) \Rightarrow A^{-1} = VD^{-1}U^T$
- $\forall \bar{y} \in A(S_2) \, \exists \bar{x} \in S_2 : \; \bar{y} = A\bar{x}$ and foe $\bar{w} = U^T \bar{y}$

$$1 = ||\bar{x}||_2^2 = ||A^{-1}A\bar{x}||_2^2 = ||A^{-1}\bar{y}||_2^2 = ||VD^{-1}U^T\bar{y}||_2^2$$

$$= ||D^{-1}U^T\bar{y}||_2^2 = ||D^{-1}\bar{w}||_2^2 = \frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} + \ldots + \frac{w_n^2}{\sigma_n^2}$$

- $U^T A(S2)$ is an ellipsoid whose kth semiaxis has length σ_k.

Condition Number of a Matrix

- U is orthogonal $\Rightarrow ||U^T A(S2)|| = ||A(S_2)||$ and $A(S_2)$ is also an ellipsoid with kth semiaxis of length σ_k.
- The ellipsoid $U^T A(S_2)$ is in standard position and its axes are directed along the standard basis vectors.
- $UU^T A(S_2) = A(S_2) \Rightarrow$The axes of $A(S_2)$ are directed along the left-hand singular vectors defined by the columns of U and the kth semiaxis of $A(S_2)$ is $\sigma_k \bar{u}_k$.
- The degree of distortion of the unit sphere under transformation by A is therefore measured by$\kappa = \frac{\sigma_1}{\sigma_n}$, the ratio of the largest singular value to the smallest singular value.

Condition Number of a Matrix

- $\max_{||\bar{x}||_2=1} ||A\bar{x}||_2 = ||A||_2 = ||UDV^T||_2 = ||D||_2 = \sigma_1$
- $\min_{||\bar{x}||_2=1} ||A\bar{x}||_2 = \frac{1}{||A^{-1}||_2} = \frac{1}{||VD^{-1}U^T||_2} = \frac{1}{||D^{-1}||_2} = \sigma_n$
- The longest and shortest vectors on $A(S_2)$ have respective lengths σ_1 and σ_n and

$$\kappa_2 = \frac{\sigma_1}{\sigma_n}$$

  is the 2-norm condition number.
- Different norms result in condition numbers with different values but with more or less the same order of magnitude as κ_2.

Condition Number of a Matrix

Definition: Let a matrix A be invertible. The condition number of A is defined by

$$\kappa(A) = ||A|| \cdot ||A^{-1}||$$

- $1 = \max_{x \neq 0} \frac{||Ix||}{||x||} = ||I|| = ||AA^{-1}|| \leq ||A|| \cdot ||A^{-1}|| = \kappa(A)$
- $\kappa(A) = ||A||_2 \cdot ||A^{-1}||_2 = \sigma_{max}(A) \frac{1}{\sigma_{min}(A)}$
- Define $M = ||A|| = \max_{x \neq 0} \frac{||Ax||}{||x||}$ and $m = \min_{x \neq 0} \frac{||Ax||}{||x||}$

$$m = \min_{x \neq 0} \frac{||Ax||}{||x||} = [y = Ax] = \min_{y \neq 0} \frac{||y||}{||A^{-1}y||} = \frac{1}{\max_{y \neq 0} \frac{||A^{-1}y||}{||y||}} = \frac{1}{||A^{-1}||}$$

- $\kappa(A) = \frac{M}{m} \Rightarrow$ if A^{-1} does not exist
  $\Rightarrow \det A = 0 \Rightarrow \exists x \neq 0 : Ax = 0 \Rightarrow m = 0 \Rightarrow \kappa(A) \to \infty$

A large condition number corresponds to the case when a matrix is close to being singular $\Rightarrow$ an ill-conditioned matrix

Ill-conditioned Matrices

- Consider a linear system $Ax = b$ and explore the effect of the error in the RHS $Ax = b + e$ in the solution.
- Let A be invertible. Then

$$x = A^{-1}(b+e) = \underbrace{A^{-1}b}_{\text{solution}} + \underbrace{A^{-1}e}_{\text{error in the solution}}$$

- Define the relative error in the solution by $\dfrac{||A^{-1}e||}{||A^{-1}b||}$ and the relative error in the RHS by $\dfrac{||e||}{|b||}$
- Consider the ratio of the relative errors:

$$\frac{\frac{||A^{-1}e||}{||A^{-1}b||}}{\frac{||e||}{|b||}} = \frac{||A^{-1}e||}{||e||} \cdot \frac{||b||}{||A^{-1}b||}$$

$$\max \frac{\frac{||A^{-1}e||}{||A^{-1}b||}}{\frac{||e||}{|b||}} = \max_{e \neq 0} \frac{||A^{-1}e||}{||e||} \max_{b \neq 0} \frac{||b||}{||A^{-1}b||} = \max_{e \neq 0} \frac{||A^{-1}e||}{||e||} \max_{y \neq 0} \frac{||Ay||}{||y||}$$

Eckort-Young-Mirsky Theorem

Question: For a given matrix $A_{n\times m}$ and a given integer number $s \geq 1$, find a matrix C with $rank(C) = s$ which is closest to A, that is, find

$$\min_{C:\, rank\, C = s} ||A - C||$$

Answer: Let $1 \leq s \leq rank(A)$, $A_{n\times m}$. The truncated SVD

$$A_s = \sum_{i=1}^{s} \sigma_i \bar{u}_i \bar{v}_i^T$$

of a matrix $A_{n\times m}$ is the best *rank s* approximation to A and

$$\min ||A - C||_F = ||A - A_s||_F = \sqrt{\sum_{i>s} \sigma_i^2}$$

$$\min ||A - C||_2 = ||A - A_s||_2 = \sigma_{s+1}$$

Example: Find the best rank-1 approximation to the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$$

Eckort-Young-Mirsky Theorem: Example

Find the best rank-1 approximation to the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$$