Linear Algebra

Chapter 7: Distance and Approximation

University of Seoul
School of Computer Science
Minho Kim

Table of contents

Outline

Outline

Outline

Outline

Outline

Diagonalization

- For symmetric matrices,

$$A = PDP^T$$

  (Orthogonal diagonalization)
- For (diagonalizable) non-symmetric matrices,

$$A = PDP^{-1}$$

- What if a matrix is not disgonalizable?
  $\rightarrow$ SVD (Singular Vector Decomposition)

$$A = PDQ^T$$

  - P & Q orthogonal
  - D "diagonal"
  - Works for *any* matrix!

The Singular Values of a Matrix

- For any $m \times n$ matrix A,
  - $A^T A \in \mathbb{R}^{n \times n}$ is symmetric hence $A^T A$ can be orthogonally diagonalizable and
  - the eigenvalues of $A^T A$ are all real and non-negative. (Why?)
    $\rightarrow$ We can take the (positive) square roots of the eigenvalues of $A^T A$.

Definition

If A is an $m \times n$ matrix, the *singular values* of A are the square roots of the eigenvalues of $A^T A$ and are denoted by $\sigma_1, \ldots, \sigma_n$. It is conventional to arrange the singular values so that $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n$.

The Singular Values of a Matrix (cont'd)

Let $\{\mathbf{v}_1, \dots \mathbf{v}_n\}$ be the orthonormal basis for $\mathbb{R}^n$ that consists of eigenvalues of $A^T A$, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$.

- The singular values of A are the lengths of the vectors $A\mathbf{v}_1, \dots, A\mathbf{v}_n$.
- For $n = 2$, σ_1 and σ_2 are the lengths of half of the major and minor axes of the ellipse in $\mathbb{R}^m$, the image of the unit circle in $\mathbb{R}^2$.

The Singular Value Decomposition

Our goal:

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

- $A \in \mathbb{R}^{m\times n}$
- $U \in \mathbb{R}^{m\times m}$ orthogonal matrix
- $V \in \mathbb{R}^{n\times n}$ orthogonal matrix
- $\Sigma \in \mathbb{R}^{m\times n}$ “diagonal” matrix

The Singular Value Decomposition (cont'd)

- For $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$,

$$\Sigma = \left[\begin{array}{c:c} D & O \\ \hdashline O & O \end{array}\right] \quad D = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix}$$

- The columns of V are the orthonormal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$

  $\rightarrow$ Always exists. Why?

- For $r \geqslant m$, the columns of U are the normalized vectors of $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\}$ (orthogonal. Why?)

$$U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} \quad \mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$$

- For $r < m$, $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$ are obtained by extending the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$. $\rightarrow$ tricky!

The Singular Value Decomposition (cont'd)

The Singular Value Decomposition

Let A be an $m \times n$ matrix with singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Then there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ matrix Σ of the form shown in equation (1) such that

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

- *left singular vectors* of A: the columns of U
- *right singular vectors* of A: the columns of V
- U and V are not uniquely determined by A
- Σ must contain the singular values of A
- What if A is positive definite (i.e. all eigenvalues are positive) and symmetric? $\rightarrow$ The spectral theorem

The Outer Product Form of the SVD

The Outer Product Form of the SVD

Let A be an $m \times n$ matrix with singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Let $\mathbf{u}_1, \dots \mathbf{u}_r$ be left singular vectors and let $\mathbf{v}_1, \dots \mathbf{v}_r$ be right singular vectors of A corresponding to these singular values. Then

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

- What if A is positive definite and symmetric? $\rightarrow$ the spectral decomposition

Information of a Matrix Contained in Its SVD

Theorem 7.15

Let $A = U\Sigma V^T$ be a singular value decomposition of an $m \times n$ matrix A. Let $\sigma_1, \ldots, \sigma_r$ be all the nonzero singular values of A. Then

- The rank of A is r.
- $\{\mathbf{u}_1, \ldots \mathbf{u}_r\}$ is an orthonormal basis for $\mathrm{col}(A)$.
- $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_,\}$ is an orthonormal basis for $\mathrm{null}(A^T)$.
- $\{\mathbf{v}_1, \ldots \mathbf{v}_r\}$ is an orthonormal basis for $\mathrm{row}(A)$.
- $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathrm{null}(A)$.

Geometric Insight of a Matrix by Its SVD

Theorem 7.16

Let $A = U\Sigma V^T$ be a singular value decomposition of an $m \times n$ matrix A with rank r. Then the image of the unit sphere in $\mathbb{R}^n$ under the matrix transformation that maps $\mathbf{x}$ to $A\mathbf{x}$ is

- the surface of an ellipsoid in $\mathbb{R}^m$ if $r = n$.
- a solid ellipsoid in $\mathbb{R}^m$ if $r < n$.

Effect of A on the unit sphere in $\mathbb{R}^n$

1. V^T: maps the unit sphere to itself.
2. Σ: collapses $n - r$ dimensions of the unit sphere, leaving an r-dimensional unit sphere, then distort it into an ellipsoid according to $\sigma_1, \ldots, \sigma_r$.
3. U: aligns the axes of the ellipsoid with the orthonormal basis vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$ in $\mathbb{R}^m$.

Application: Matrix Approximation & Image Compression

- The matrix $A \in \mathbb{R}^{m \times n}$ requires mn storage.
- The SVD of A in outer product form

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

- If we keep only up to $k \leqslant r$ terms,

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

  A_k requires $k + km + kn = k(m + n + 1)$ storage.
- A_k is the best k-rank least-square approximation (measured by Frobenius norm) of A. In other words, the error

$$\sum_{i=1}^{m} \sum_{j=1}^{n} (A(i,j) - A_k(i,j))^2$$

  is minimized.

Outline