CSE 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

Lecture 5

Overview

- Matrix properties via singular value decomposition (SVD)
- Geometric interpretation of SVD
- Applications

Reading

- Chapter 3 of *Matrix Computations* by Gene Golub and Charles Van Loan

Matrix multiplication

- Recall

$$\mathbf{x} \otimes \mathbf{y} = \mathbf{x}\mathbf{y}^\top$$

- Let $A = [\mathbf{a}_1 \ldots \mathbf{a}_n] \in \mathbb{R}^{m \times n}$ and $B = \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_n^\top \end{bmatrix} \in \mathbb{R}^{n \times q}$, matrix multiplication can be written as

$$AB = \sum_{i=1}^{n} \mathbf{a}_i \mathbf{b}_i^\top = \sum_{i=1}^{n} \mathbf{a}_i \otimes \mathbf{b}_i$$

- For example,

$$\begin{bmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 3 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix}$$
$$+ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ -3 & -1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 6 & 3 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 4 & 2 \end{bmatrix}$$

SVD expansion

- We can decompose A in terms of singular values and vectors.

$$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i$$

  where $\otimes$ is the Kronecker product.

- The matrix 2-norm and Frobenius norm properties have connections to the SVD.

$$\begin{aligned} \|A\|_F &= \sqrt{\sigma_1^2 + \cdots + \sigma_p^2}, \quad p = \min(m, n) \\ \|A\|_2 &= \sigma_1 \\ \min_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} &= \sigma_n, \quad m \geq n \end{aligned}$$

- Closely related to eigenvalues, eigenvectors and principal component analysis.

Matrix properties via SVD

Theorem

The rank of A is r, the number of nonzero singular values.

Proof.

The rank of a diagonal matrix is equal to the number of its nonzero entries, and in SVD, $A = U\Sigma V^\top$ where U and V are of full rank. Thus, $\text{rank}(A) = \text{rank}(\Sigma) = r$. □

Theorem

$\|A\|_2 = \sigma_1$, *and* $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$

Proof.

As U and V are orthogonal, $A = U\Sigma V^\top$, $\|A\|_2 = \|\Sigma\|_2$. By definition, $\|\Sigma\|_2 = \max_{\|\mathbf{x}\|=1} \|\Sigma\mathbf{x}\|_2 = \max\{|\sigma_i|\} = \sigma_1$. Likewise, $\|A\|_F = \|\Sigma\|_F$, and by definition $\|\Sigma\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$ □

Matrix properties via SVD (cont'd)

Theorem

The nonzero singular values of A are the square roots of the nonzero eigenvalues of $AA^\top$ or $A^\top A$ (they have the same nonzero eigenvalues.).

Proof.

From definition,

$$AA^\top = (U\Sigma V^\top)(U\Sigma V^\top)^\top = U\Sigma V^\top V\Sigma U^\top = U \text{ diag}(\sigma_1^2, \ldots, \sigma_p^2) \; U^\top$$

$\square$

Theorem

For $A \in \mathbb{R}^{m\times m}$, $|det(A)| = \prod_{i=1}^m \sigma_i$

Proof.

$$|\det(A)| = |\det(U\Sigma V^\top)| = |\det(U)||\det(\Sigma)||\det(V^\top)| = |\det(\Sigma)| = \prod_{i=1}^m \sigma_i$$

$\square$

Matrix properties via SVD (cont'd)

Theorem

A is the sum of r rank one matrices: $A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_j^\top$

Theorem

(Eckart-Young 1936) Let $A = U\Sigma V^\top = U\ \mathit{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) V^\top$.
For any ν *with* $0 \le \nu \le r$, $A_\nu = \sum_{i=1}^{\nu} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$,

$$\|A - A_\nu\|_2 = \min_{\mathit{rank}(B) \le \nu} \|A - B\|_2 = \sigma_{\nu+1}$$

Proof.

Let $\Sigma_\nu = U(A - A_\nu)V^\top$, then

$$\begin{aligned} \Sigma_\nu &= U\ (\mathrm{diag}(\sigma_1, \ldots, \sigma_\nu, \sigma_{\nu+1}, \ldots, \sigma_p) - \mathrm{diag}(\sigma_1, \ldots, \sigma_\nu, 0, \ldots, 0)) V^\top \\ &= U\ \mathrm{diag}(0, \ldots, 0, \sigma_{\nu+1}, \ldots, \sigma_p)\ V^\top \end{aligned}$$

, consequently $\|A - A_\nu\|_2 = \|\Sigma_\nu\|_2 = \sigma_{\nu+1}$. $\square$

Theorem

(Eckart-Young 1936) Let $A = U\Sigma V^\top = U \; diag(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) V^\top$*.*
For any ν *with* $0 \leq \nu \leq r$, $A_\nu = \sum_{i=1}^{\nu} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$,

$$\|A - A_\nu\|_2 = \min_{rank(B) \leq \nu} \|A - B\|_2 = \sigma_{\nu+1}$$

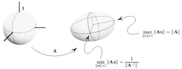
Proof.

Suppose there is some B with $\text{rank}(B) \leq \nu$ such that $\|A - B\|_2 < \|A - A_\nu\|_2 = \sigma_{\nu+1}$. Then there exists an $n - \nu$ dimensional subspace $W \in \mathbb{R}^n$ such that $\mathbf{w} \in W \Rightarrow B\mathbf{w} = 0$. Then

$$\|A\mathbf{w}\|_2 = \|(A - B)\mathbf{w}\|_2 \leq \|A - B\|_2 \|\mathbf{w}\|_2 \leq \sigma_{\nu+1} \|\mathbf{w}\|_2$$

Thus W is a $n - \nu$ dimensional subspace where $\|A\mathbf{w}\| < \sigma_{\nu+1}\|\mathbf{w}\|$. But there is a $\nu + 1$ dimensional subspace where $\|A\mathbf{w}\| \geq \sigma_{\nu+1}\|\mathbf{w}\|$, namely the space spanned by the first $\nu + 1$ right singular vector of A. Since the sum of the dimensions of these two spaces exceeds n, there must be a nonzero vector lying in both, and this is a contraction. □

Geometric interpretation of Eckart-Young theorem



- What is the best approximation of a hyperellipsoid by a line segment?
  - Take the line segment to be the longest axis.
  - Next, what is the best approximation by a two-dimensional ellipsoid?
  - Take the ellipsoid spanned by the longest and the second longest axis.
  - Continue this fashion, at each step we improve the approximation by adding into our approximation the largest axis of the hyperellipsoid not yet included.
- Reminiscent of techniques used in image compression, machine learning, and functional analysis (e.g., matching pursuit).

Theorem

For any ν with $0 \leq v \leq r$, $A_\nu = \sum_{i=1}^{\nu} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$,

$$\|A - A_\nu\|_F = \min_{rank(B)\leq\nu} = \sqrt{\sigma_{\nu+1}^2 + \cdots + \sigma_r^2}$$