Matrix methods – Singular value decomposition

MATH 2740 – Mathematics of Data Science – Lecture 09

Julien Arino

julien.arino@umanitoba.ca

Department of Mathematics @ University of Manitoba

Fall 202X

The University of Manitoba campuses are located on original lands of Anishinaabeg, Ininew, Anisininew, Dakota and Dene peoples, and on the National Homeland of the Red River Métis. We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of Reconciliation and collaboration.

Outline

Computing the SVD

Applications of the SVD – Least squares

Summary – Least squares methods

Computing the SVD

Applications of the SVD – Least squares

Summary – Least squares methods

Computing the SVD (case of $\neq$ eigenvalues)

To compute the SVD, we use the following result

Theorem 7

Let $A \in \mathcal{M}_n$ symmetric, $(\lambda_1, \boldsymbol{u}_1)$ and $(\lambda_2, \boldsymbol{u}_2)$ be eigenpairs, $\lambda_1 \neq \lambda_2$. Then $\boldsymbol{u}_1 \bullet \boldsymbol{u}_2 = 0$

Proof of Theorem 7

$A \in \mathcal{M}_n$ symmetric, $(\lambda_1, \boldsymbol{u}_1)$ and $(\lambda_2, \boldsymbol{u}_2)$ eigenpairs with $\lambda_1 \neq \lambda_2$

$$
\begin{aligned}
\lambda_1(\boldsymbol{v}_1 \bullet \boldsymbol{v}_2) &= (\lambda_1 \boldsymbol{v}_1) \bullet \boldsymbol{v}_2 \\
&= A\boldsymbol{v}_1 \bullet \boldsymbol{v}_2 \\
&= (A\boldsymbol{v}_1)^T \boldsymbol{v}_2 \\
&= \boldsymbol{v}_1^T A^T \boldsymbol{v}_2 \\
&= \boldsymbol{v}_1^T (A\boldsymbol{v}_2) \qquad [A \text{ symmetric so } A^T = A] \\
&= \boldsymbol{v}_1^T (\lambda_2 \boldsymbol{v}_2) \\
&= \lambda_2(\boldsymbol{v}_1^T \boldsymbol{v}_2) \\
&= \lambda_2(\boldsymbol{v}_1 \bullet \boldsymbol{v}_2)
\end{aligned}
$$

So $(\lambda_1 - \lambda_2)(\boldsymbol{v}_1 \bullet \boldsymbol{v}_2) = 0$. But $\lambda_1 \neq \lambda_2$, so $\boldsymbol{v}_1 \bullet \boldsymbol{v}_2 = 0$ $\square$

Computing the SVD (case of $\neq$ eigenvalues)

If all eigenvalues of A^TA (or AA^T) are distinct, we can use Theorem 7

1. Compute $A^TA \in \mathcal{M}_n$
2. Compute eigenvalues $\lambda_1, \ldots, \lambda_n$ of A^TA; order them as $\lambda_1 > \cdots > \lambda_n \geq 0$ ($>$ not $\geq$ since $\neq$)
3. Compute singular values $\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_n = \sqrt{\lambda_n}$
4. Diagonal matrix D in Σ is either in $\mathcal{M}_n$ (if $\sigma_n > 0$) or in $\mathcal{M}_{n-1}$ (if $\sigma_n = 0$)

5. Since eigenvalues are distinct, Theorem 7 $\implies$ eigenvectors are orthogonal set. Compute these eigenvectors in the same order as the eigenvalues
6. Normalise them and use them to make the matrix V, i.e., $V = [\boldsymbol{v}_1 \cdots \boldsymbol{v}_n]$
7. To find the $\boldsymbol{u}_i$, compute, for $i = 1, \ldots, r$,

$$\boldsymbol{u}_i = \frac{1}{\sigma_i} A\boldsymbol{v}_i$$

and ensure that $\|\boldsymbol{u}_i\| = 1$

Computing the SVD (case where some eigenvalues are $=$)

1. Compute $A^T A \in \mathcal{M}_n$
2. Compute eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A^T A$; order them as $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$
3. Compute singular values $\sigma_1 = \sqrt{\lambda_1}, \ldots, \sigma_n = \sqrt{\lambda_n}$, with $r \leq n$ the index of the last positive singular value
4. For eigenvalues that are distinct, proceed as before
5. For eigenvalues with multiplicity > 1, we need to ensure that the resulting eigenvectors are LI *and* orthogonal

Dealing with eigenvalues with multiplicity > 1

When an eigenvalue has (algebraic) multiplicity > 1, e.g., characteristic polynomial contains a factor like $(\lambda - 2)^2$, things can become a little bit more complicated

The proper way to deal with this involves the so-called Jordan Normal Form (another matrix decomposition)

In short: not all square matrices are diagonalisable, but all square matrices admit a JNF

Sometimes, we can find several LI eigenvectors associated to the same eigenvalue. Check this. If not, need to use the following

Definition 8 (Generalised eigenvectors)

The vector $\boldsymbol{x} \neq \mathbf{0}$ is a **generalized eigenvector** of rank m of $A \in \mathcal{M}_n$ corresponding to eigenvalue λ if

$$(A - \lambda \mathbb{I})^m \boldsymbol{x} = \mathbf{0}$$

but

$$(A - \lambda \mathbb{I})^{m-1} \boldsymbol{x} \neq \mathbf{0}$$

Procedure for generalised eigenvectors

$A \in \mathcal{M}_n$ and assume λ eigenvalue with algebraic multiplicity k

Find $\boldsymbol{v}_1$, "classic" eigenvector, i.e., $\boldsymbol{v}_1 \neq \boldsymbol{0}$ s.t. $(A - \lambda\mathbb{I})\boldsymbol{v}_1 = \boldsymbol{0}$

Find generalised eigenvector $\boldsymbol{v}_2$ of rank 2 by solving for $\boldsymbol{v}_2 \neq \boldsymbol{0}$,

$$(A - \lambda\mathbb{I})\boldsymbol{v}_2 = \boldsymbol{v}_1$$

$\ldots$

Find generalised eigenvector $\boldsymbol{v}_k$ of rank k by solving for $\boldsymbol{v}_k \neq \boldsymbol{0}$,

$$(A - \lambda\mathbb{I})\boldsymbol{v}_k = \boldsymbol{v}_{k-1}$$

Then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ LI

Back to the normal procedure

With the LI eigenvectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ corresponding to λ

Apply Gram-Schmidt to get orthogonal set

For all eigenvalues with multiplicity > 1, check that you either have LI eigenvectors or do what we just did

When you are done, be back on your merry way to step 6 in the case where eigenvalues are all $\neq$

I am caricaturing a little here: there can be cases that do not work exactly like this, but this is general enough..

Pseudoinverse of a matrix

Definition 9 (Pseudoinverse)

$A = U\Sigma V^T$ an SVD for $A \in \mathcal{M}_{mn}$, where

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \text{ with } D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$$

(D contains the nonzero singular values of A ordered as usual)

The **pseudoinverse** (or **Moore-Penrose inverse**) of A is $A^+ \in \mathcal{M}_{nm}$ given by

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

with

$$\Sigma^+ = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}_{nm}$$

Least squares revisited

Theorem 10

Let $A \in \mathcal{M}_{mn}$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^m$. The least squares problem $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least squares solution $\tilde{\boldsymbol{x}}$ of minimal length (closest to the origin) given by

$$\tilde{\boldsymbol{x}} = A^+\boldsymbol{b}$$

where A^+ is the pseudoinverse of A

Computing the SVD

Applications of the SVD – Least squares

Summary – Least squares methods

The least squares problem

Problem Statement:

Given a system $A\boldsymbol{x} = \boldsymbol{b}$ where $A \in \mathcal{M}_{mn}$, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$ (typically $m > n$), find $\tilde{\boldsymbol{x}}$ that minimizes

$$\|\boldsymbol{b} - A\boldsymbol{x}\|^2 = \sum_{i=1}^{m}(b_i - \sum_{j=1}^{n} A_{ij}x_j)^2$$

Geometric interpretation: Find the vector $A\tilde{\boldsymbol{x}}$ in the column space of A that is closest to $\boldsymbol{b}$

Solution: $A\tilde{\boldsymbol{x}} = \text{proj}_{\text{col}(A)}(\boldsymbol{b})$

Method 1: Normal equations

The normal equations:

$$A^T A\tilde{\boldsymbol{x}} = A^T \boldsymbol{b}$$

When this works:

- Always has at least one solution
- Any solution $\tilde{\boldsymbol{x}}$ to the normal equations is a least squares solution

Computational issues:

- Forming $A^T A$ can be numerically unstable
- Condition number of $A^T A$ is the square of the condition number of A
- Still useful for theoretical analysis

Method 2: when A Has linearly independent columns

Condition: $A \in \mathcal{M}_{mn}$ has linearly independent columns

Then: $A^T A$ is invertible and the least squares solution is **unique**

$$\tilde{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$$

Properties:

- $A^T A \in \mathcal{M}_n$ is square, symmetric, and positive definite
- $(A^T A)^{-1} A^T$ is called the *left pseudoinverse* of A
- This gives the unique least squares solution

Drawback: Computing $(A^T A)^{-1}$ directly can be numerically unstable

Method 3: QR factorization

QR Factorization: If $A \in \mathcal{M}_{mn}$ has linearly independent columns, then

$$A = QR$$

where $Q \in \mathcal{M}_{mn}$ has orthonormal columns and $R \in \mathcal{M}_n$ is upper triangular and nonsingular

Least squares solution:

$$\tilde{\boldsymbol{x}} = R^{-1} Q^T \boldsymbol{b}$$

Advantages:

- More numerically stable than forming $A^T A$
- R is upper triangular $\Rightarrow$ solving $R\tilde{\boldsymbol{x}} = Q^T \boldsymbol{b}$ by back substitution
- Condition number of R equals condition number of A
- Gram-Schmidt or Householder reflections can compute QR factorization

Method 4: Singular Value Decomposition (SVD)

SVD: For any $A \in \mathcal{M}_{mn}$,

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

where $U \in \mathcal{M}_m$ orthogonal, $V \in \mathcal{M}_n$ orthogonal, $\Sigma \in \mathcal{M}_{mn}$ with $\Sigma_{ii} = \sigma_i \geq 0$ (singular values)

Pseudoinverse: $A^+ = V\Sigma^+ U^T$ where Σ^+ has $(\Sigma^+)_{ii} = 1/\sigma_i$ if $\sigma_i > 0$, else 0

Least squares solution:

$$\tilde{\boldsymbol{x}} = A^+ \boldsymbol{b}$$

Key advantages:

- Works for *any* matrix A (even when columns are linearly dependent)
- Gives the solution of *minimal length* when multiple solutions exist
- Most numerically stable method
- Reveals the rank of A through the number of non-zero singular values

When to use which method

Method	**When to use**	**Advantages/Drawbacks**
Normal equations	Theory, small problems	Simple, but unstable
$(A^T A)^{-1} A^T$	A has LI columns	Explicit formula, unstable
QR Factorization	A has LI columns	Stable, efficient
SVD	Any A, rank-deficient	Most stable, handles all cases

Use QR for well-conditioned problems with LI columns, SVD for rank-deficient or ill-conditioned problems