Mathematics for Computer Science
Linear Algebra

Lecture 19: Singular value decomposition

Andrei Krokhin

March 14, 2021

Matrix decompositions we've seen so far

- LU-decomposition $A = LU$, L is lower triangular, U is upper triangular
  - Condition: A square, all principal minors non-0
- PLU-decomposition $A = PLU$, P is a permutation matrix, L and U as above
  - Condition: A square
- QR-decomposition $A = QR$, Q has orthonormal columns, R is invertible upper triangular
  - Condition: A has linearly independent columns
- Eigendecomposition $A = PDP^{-1}$, D is diagonal, P is invertible
  - Condition: A has size $n \times n$ and n linearly independent eigenvectors
- Spectral decomposition $A = QDQ^T$, D is diagonal, Q is orthogonal
  - Condition: A is symmetric (equivalently, has n orthonormal eigenvectors)

A and $A^T A$

Theorem

For any $m \times n$ matrix A, the following holds:

1. *A and $A^T A$ have the same null space.*
2. *A and $A^T A$ have the same row space.*
3. *A and $A^T A$ have the same rank.*

Proof.

We proved item (1) in lecture 16 (about least squares).

(1) implies (2), since row space is the orthogonal complement of the null space.

(3) follows immediately from (2), since rank is the dimension of the row space. $\square$

Eigenvalues of $A^T A$

Theorem

For any $m \times n$ matrix A, the eigenvalues of $A^T A$ are non-negative.

(Symmetric matrices whose eigenvalues are all non-negative are called *positive semidefinite*.)

Proof.

Since $A^T A$ is symmetric, the spectral theorem says that $\mathbb{R}^n$ has an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ consisting of eigenvectors of $A^T A$.

Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues. Then, for any $1 \leq i \leq n$, we have

$$||A\mathbf{v}_i||^2 = \langle A\mathbf{v}_i, A\mathbf{v}_i \rangle = (A\mathbf{v}_i)^T A\mathbf{v}_i = \mathbf{v}_i^T A^T A\mathbf{v}_i = \mathbf{v}_i^T(\lambda_i \mathbf{v}_i) = \lambda_i(\mathbf{v}_i^T \mathbf{v}_i) = \lambda_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle = \lambda_i ||\mathbf{v}_i||^2 = \lambda_i.$$

Thus, $\lambda_i = ||A\mathbf{v}_i||^2 \geq 0$. $\square$

Singular values

Definition

If A is an $m \times n$ matrix and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of $A^T A$ then the singular values of A are the numbers

$$\sigma_1 = \sqrt{\lambda_1},\ \sigma_2 = \sqrt{\lambda_2},\ \ldots\ ,\ \sigma_n = \sqrt{\lambda_n}.$$

Example: Let

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

Then

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

The eigenvalues of $A^T A$ are $\lambda_1 = 3$ and $\lambda_2 = 1$, so the singular values of A are

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3} \ \text{ and } \ \sigma_2 = \sqrt{\lambda_2} = 1.$$

Full singular value decomposition (SVD)

Theorem

If A is an $m \times n$ matrix of rank k then A can be decomposed as $A = U\Sigma V^T =$

$$(\mathbf{u}_1|\ldots|\mathbf{u}_k|\ldots|\mathbf{u}_m)\left(\begin{array}{cccc|c} \sigma_1 & 0 & \ldots & 0 & \\ 0 & \sigma_2 & \ldots & 0 & \\ \vdots & \vdots & \ddots & \vdots & O_{k\times(n-k)} \\ 0 & 0 & \ldots & \sigma_k & \\ \hline & & O_{(m-k)\times k} & & O_{(m-k)\times(n-k)} \end{array}\right)\left(\begin{array}{c} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \vdots \\ \mathbf{v}_n^T \end{array}\right)$$

where U, Σ, and V have sizes $m \times m$, $m \times n$ and $n \times n$, respectively, and

1. $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$ *are the non-zero singular values of A.*
2. $V = (\mathbf{v}_1|\ldots|\mathbf{v}_k|\ldots|\mathbf{v}_n)$ *is orthogonal, it orthogonally diagonalises A^TA.*
3. $U = (\mathbf{u}_1|\ldots|\mathbf{u}_k|\ldots|\mathbf{u}_m)$ *is orthogonal, it orthogonally diagonalises AA^T.*
4. $\mathbf{u}_i = \frac{A\mathbf{v}_i}{||A\mathbf{v}_i||} = \frac{1}{\sigma_i}A\mathbf{v}_i$ *for $i = 1, \ldots, k$.*

Remarks

- The number k of non-0 singular values $(\sigma_1, \ldots, \sigma_k)$ is equal to the rank of A.
- The columns of V are orthonormal eigenvectors of $A^T A$, with $\mathbf{v}_1, \ldots, \mathbf{v}_k$ ordered so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$.
- Vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are called the left singular vectors of A.
  - They form an orthonormal basis for the column space of A.
- Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are called the right singular vectors of A.
  - They form an orthonormal basis for the row space of A.
- Vectors $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ form an orthonormal basis for the null space of A.
- Vectors $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m$ form an orthonormal basis for the null space of A^T.

- To compute SVD, first orthogonally diagonalise $A^T A$ - this gives V and Σ. Then find $\mathbf{u}_1, \ldots, \mathbf{u}_k$ as in item (3) in the theorem, and then extend this set to an orthonormal basis of $\mathbb{R}^m$ to complete U.
- There are other algorithms to compute SVD (or its most important parts)

Example

Let

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

We already found eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 1$ of $A^T A$, and $\sigma_1 = \sqrt{3}, \sigma_2 = 1$.

The corresponding eigenvectors of $A^T A$ are $\mathbf{v}_1 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ and $\mathbf{v}_2 = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$.

Now compute $\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = (\frac{\sqrt{6}}{3}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6})$ and $\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = (0, -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$.

To extend $\{\mathbf{u}_1, \mathbf{u}_2\}$ to an orthonormal basis of $\mathbb{R}^3$, can find an orthonormal basis in $W^\perp$ where $W = \mathit{span}(\mathbf{u}_1, \mathbf{u}_2)$. One such basis is $\{\mathbf{u}_3 = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})\}$.

Thus, one singular value decomposition of A is

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

(Sketch of) the proof of the SVD theorem

Choose $V = [\mathbf{v}_1 | \ldots | \mathbf{v}_n]$ so that it orthogonally diagonalises $A^T A$ - this means that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthonormal set of eigenvectors of $A^T A$.

If needed, order that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ so that the corresponding eigenvalues of $A^T A$ satisfy $\lambda_1 \geq \ldots \geq \lambda_n$. We proved earlier (slide 4) that $||A\mathbf{v}_i||^2 = \lambda_i$.

Since $rank(A^T A) = rank(A) = k$, we have $\lambda_1 \geq \ldots \geq \lambda_k > \lambda_{k+1} = \ldots = \lambda_n = 0$.

Let σ_i's and $\mathbf{u}_i$'s $(1 \leq i \leq k)$ be as in the theorem. Then

- we have $\sigma_i \mathbf{u}_i = A\mathbf{v}_i$ and $||\mathbf{u}_i|| = 1$ $(1 \leq i \leq k)$ by the choice of $\mathbf{u}_i$'s
- we have $A\mathbf{v}_i = \mathbf{0}$ $(k+1 \leq i \leq n)$ because $||A\mathbf{v}_i||^2 = \lambda_i = 0$ for $i \geq k+1$.

It follows that $A = U\Sigma V^T$, since this is equivalent to

$$U\Sigma = [\sigma_1 \mathbf{u}_1 | \ldots | \sigma_k \mathbf{u}_k | \ \mathbf{0} \ | \ldots | \ \mathbf{0} \ ] = [A\mathbf{v}_1 | \ldots | A\mathbf{v}_k | \ldots | A\mathbf{v}_n] = AV.$$

Finally, show that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal basis in the column space of A, and if we add to it any orthonormal basis $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ for the null space of A^T then $U = (\mathbf{u}_1 | \ldots | \mathbf{u}_k | \ldots | \mathbf{u}_m)$ orthogonally diagonalises AA^T (in the practical!).

Reduced SVD

The matrix Σ in full SVD has three all-0 submatrices. We can get rid of them:

$$A = U_k \Sigma_k V_k^T = (\mathbf{u}_1 | \ldots | \mathbf{u}_k) \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_k \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \end{pmatrix}.$$

This is called a reduced singular value decomposition of A.

Here U_k, Σ_k, and V_k and have sizes $m \times k$, $k \times k$ and $k \times n$, respectively.

Note that the diagonal elements of Σ_k are all positive, so Σ_k is invertible.

Multiplying out matrices in the reduced SVD above, we get

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \ldots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T,$$

which is called a reduced singular value expansion of A.

Example

From an earlier example, we have a singular value decomposition $A = U\Sigma V^T$

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

Its reduced form $A = U_2\Sigma_2 V_2^T$ is

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

The corresponding singular value expansion $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T$ is

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

(In different sources, SVD can mean any of these three forms).

Application: Rank-r approximation

- For an $m \times n$ matrix A of rank k, its SVD expansion $A = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ consists of $k(1+m+n)$ numbers, which can be much smaller than mn.
- For $r \leq k$, the following matrix A_r is called the rank-r approximation of A

$$A_r = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \ldots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T.$$

If $\sigma_1, \ldots, \sigma_r$ are much larger than $\sigma_{r+1}, \ldots, \sigma_k$ then A_r can be thought of as the "core data" in A, while the rest (i.e. $A - A_r$) is the "noise".

Theorem (Eckart-Young theorem)

For any $m \times n$ matrix A, its rank-r approximation A_r has rank r and we have

$$||A - B||_F \geq ||A - A_r||_F = \sqrt{\sigma_{r+1}^2 + \ldots + \sigma_k^2}$$

for all $m \times n$ matrices B of rank at most r.

$||...||_F$ is the Frobenius norm of a matrix: if $X = (x_{ij})$ then $||X||_F = \sqrt{\sum_{i,j} x_{ij}^2}$

Low rank approximation in image compression

The images are created by using the `svd` class of Python's `numpy.linalg` module

Original image by Jonathan McIntosh, licensed under Creative Commons.

Stability: Singular values vs. eigenvalues

Consider two square matrices:

$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ \frac{1}{60000} & 0 & 0 & 0 \end{pmatrix}$
Eigenvalues: $\lambda = 0, 0, 0, 0$	Eigenvalues: $\lambda = \frac{1}{10}, -\frac{1}{10}, \frac{i}{10}, -\frac{i}{10}$
Singular values: $\sigma = 3, 2, 1$	Singular values: $\sigma = 3, 2, 1, \frac{1}{60000}$
Rank $= 3$	Rank $= 4$

- A small change in a matrix can significantly change eigenvalues.
- Singular values of any matrix are stable: they don't change more than we change the matrix. They can determine the "effective" rank of a matrix.

What we learnt today

- Singular value decomposition, in three forms
- Application: Rank-r approximation

Next time:

- General linear transformations