Singular Value Decomposition

Jaskirat Kaur

February 23, 2024

Diagonalization Review

Diagonalization is the process of transforming a square matrix into a special type of matrix called a diagonal matrix, in which all off-diagonal entries are zero.

$$A^T = (X)\Lambda X^{-1}$$

Here, A is the original matrix, Λ is the diagonal matrix, and X is the matrix whose columns are the eigenvectors of A.

Why Do We Care About Diagonalization?

Diagonalization is important because:

- It simplifies matrix computations, such as raising a matrix to a power.

Diagonalization and Matrix Powers

To raise A to the power of k, we use the diagonalization:

$$A^k = (PDP^{-1})^k$$

Expanding this expression:

$$A^k = PDP^{-1}PDP^{-1} \cdots PDP^{-1}$$

$$A^k = PD^kP^{-1}$$

Introduction to SVD

Singular Value Decomposition (SVD) is a method of decomposing any $m \times n$ matrix, regardless of its specific properties like squareness or invertibility. The theorem states:

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$$

where U is an orthogonal $m \times r$ matrix, Σ is an $r \times r$ diagonal matrix, and V is an orthogonal $n \times r$ matrix.

Components of SVD

Let A be an $m \times n$ matrix of rank r. Then

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T,$$

where:

- U is an orthogonal $m \times m$ matrix,
- Σ is an $m \times n$ diagonal matrix,
- V is an orthogonal $n \times n$ matrix.

Such that:

1. The first r columns of U are an orthonormal basis for the column space of A.
2. The first r columns of V are an orthonormal basis for the row space of A.
3. The last $m - r$ columns of U are an orthonormal basis for the left null space of A.
4. The last $n - r$ columns of V are an orthonormal basis for the null space of A.
5. The first r diagonal entries of Σ are the nonzero singular values of A.

SVD and Low-Rank Approximations

SVD is particularly useful for constructing low-rank approximations of A by considering only the largest singular values and corresponding singular vectors. This approach is fundamental in data compression and noise reduction.

$$A \approx \sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$$

where $k < r$ is chosen based on the desired approximation accuracy.

Image Reconstruction Using SVD

SVD Illustration

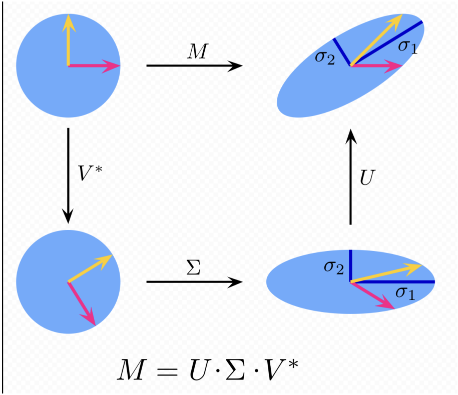

Figure: Source: Wikipedia

Singular Value Decomposition (SVD)

$$A = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}_{2\times 3} = \quad U = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix}_{2\times 2} \times \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix}_{2\times 3} \times \quad V^T = \begin{bmatrix} - & v_1^T & - \\ - & v_2^T & - \\ - & v_3^T & - \end{bmatrix}_{3\times 3}$$

Calculating the SVD

To calculate the SVD of a matrix A:

1. Compute the eigenvalues and eigenvectors of $A^T A$ and AA^T.
2. The eigenvectors of $A^T A$ form V, and the eigenvectors of AA^T form U.
3. The non-zero square roots of the eigenvalues of $A^T A$ (or AA^T) are the singular values σ_i.
4. Construct Σ by placing σ_i along the diagonal.
5. Normalize the singular vectors to ensure orthonormality.

Warm up

Singular value decomposition Q

Given the matrix $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$, give the *reduced* singular value decomposition of A.

Note. Blanks will be interpreted as 0.

$U = \begin{bmatrix} \quad & \quad \\ \quad & \quad \\ \quad & \quad \end{bmatrix}$?

$\Sigma = \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}$?

$V = \begin{bmatrix} \quad & \quad \\ \quad & \quad \\ \quad & \quad \\ \quad & \quad \\ \quad & \quad \end{bmatrix}$?

Warm up

Singular value decomposition Q

Given the matrix $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$, give the *reduced* singular value decomposition of A.

Note. Blanks will be interpreted as 0.

$$U = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

? ✓ 100%

$$\Sigma = \begin{bmatrix} 1.414 & \\ & 1 \end{bmatrix}$$

? ✓ 100%

$$V = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0.707 & 0 \\ 0.707 & 0 \end{bmatrix}$$

? ✓ 100%

Permutation Decomposition

Permutation decomposition

Consider the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$.

Is A invertible? Orthogonal? Diagonalizable? Can A be decomposed into LU? QR? $X\Lambda X^{-1}$? $Q\Lambda Q^T$? Justify your answer.

Permutation Decomposition

Permutation decomposition

Consider the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$.

Is A invertible? Orthogonal? Diagonalizable? Can A be decomposed into LU? QR? $X\Lambda X^{-1}$? $Q\Lambda Q^T$? Justify your answer.

Approach:

- Recall the definitions and properties of matrix operations and decompositions.
- Consider the characteristics of A in terms of its columns' independence, orthogonality, and symmetry.

Permutation Decomposition

- A is **invertible** since it has linearly independent columns.
- A is **orthogonal** because its columns are all unit length and orthogonal to each other, indicating orthonormal columns.
- A is **diagonalizable** since it is symmetric.

Permutation Decomposition

- A can be decomposed into **LU** because it is invertible. This property allows for such a decomposition when there are no zero pivots.
- A can be decomposed into **QR** because it has linearly independent columns. QR decomposition is applicable to matrices with full column rank.
- A can be decomposed into $X\Lambda X^{-1}$ because it is symmetric. Symmetric matrices are always diagonalizable in this form, where X can be chosen as the matrix of orthonormal eigenvectors.
- A can be decomposed into $Q\Lambda Q^T$ because it is symmetric. This is another form of diagonalization specific to symmetric matrices.
- A can be decomposed into $U\Sigma V^T$ because every matrix can be decomposed using Singular Value Decomposition (SVD), regardless of its properties.

Pseudoinverses

Pseudoinverses

Let A be a matrix with SVD $A = U\Sigma V^T$. We define the **pseudoinverse** of A to be the matrix

$$A^+ = V\Sigma^{-1}U^T,$$

where Σ^{-1} is a diagonal matrix of inverses of singular values, $\frac{1}{\sigma_i}$.

We want to prove the following:

If $\mathbf{x}^* = A^+\mathbf{b}$, then $A\mathbf{x}^*$ is the closest vector to $\mathbf{b}$ in the column space of A.

There are two possible ways to approach this. We will consider both.

1. If A^+ is the "pseudoinverse", what happens when we take AA^+? Use this to see if $A\mathbf{x}^*$ is the projection of $\mathbf{b}$ onto the column space of A.
2. If $A\mathbf{x}^*$ is the projection of $\mathbf{b}$ on the column space of A, this implies that $\mathbf{x}^* = \hat{\mathbf{x}}$. See whether this is the case by using the definition of $\hat{\mathbf{x}}$.

Here is a final question to consider. The consequence of this result is it gives us a way to compute least squares approximations without the requirement that the columns of A are linearly independent. Why?

Pseudoinverses

Approach:

- Understand the concept of pseudoinverse and how it relates to projections.
- Analyze the algebraic properties of A^+ and its effect when applied to b.

Computing AA^+

Given: $A^+ = V\Sigma^{-1}U^T$

$$\begin{aligned} AA^+ &= (U\Sigma V^T)(V\Sigma^{-1}U^T) \\ &= U(\Sigma V^T V\Sigma^{-1})U^T \\ &= U(\Sigma I\Sigma^{-1})U^T \\ &= UU^T \end{aligned}$$

Observation: $V^TV = I$ because V is orthogonal, and $\Sigma\Sigma^{-1} = I$ as Σ is diagonal. Thus, $AA^+ = UU^T$ is a projection matrix onto the column space of A.

Projection of b onto Column Space of A

Given: $x^* = A^+ b$

$$
\begin{aligned}
Ax^* &= A(A^+ b) \\
&= (AA^+)b \\
&= (UU^T)b
\end{aligned}
$$

Conclusion: $Ax^* = UU^T b$ is the projection of b onto the column space of A, since UU^T is a projection matrix.

Least Squares Solution and SVD

- Starting with the least squares solution: $\hat{x} = (A^T A)^{-1} A^T b$
- Substituting $A = U \Sigma V^T$ gives $\hat{x} = (V \Sigma^2 V^T)^{-1} (V \Sigma U^T) b$
- Simplifying, we get $\hat{x} = V \Sigma^{-2} V^T (V \Sigma U^T) b = V \Sigma^{-1} U^T b = A^+ b$

Conclusion: $\hat{x} = x^*$, showing that the pseudoinverse A^+ can be used to compute the least squares solution, $\hat{x}$, even when A does not have linearly independent columns, as every matrix has an SVD.

Left Eigenvectors

Left eigenvectors

Let A be an $n \times n$ matrix that is diagonalizable into $A = X\Lambda X^{-1}$. We define the eigenvectors of A^T to be the **left eigenvectors** of A, since $A^T\mathbf{y} = \lambda\mathbf{y}$ gives $\mathbf{y}^T A = \lambda\mathbf{y}^T$.

Show that $A = \lambda_1\mathbf{x}_1\mathbf{y}_1^T + \cdots + \lambda_n\mathbf{x}_n\mathbf{y}_n^T$, where $\mathbf{x}_i$ is the eigenvector associated with eigenvalue λ_i and $\mathbf{y}_i$ is the left eigenvector associated with eigenvalue λ_i.

Hint. What is A^T?

Left Eigenvectors

Left eigenvectors

Let A be an $n \times n$ matrix that is diagonalizable into $A = X\Lambda X^{-1}$. We define the eigenvectors of A^T to be the **left eigenvectors** of A, since $A^T\mathbf{y} = \lambda\mathbf{y}$ gives $\mathbf{y}^T A = \lambda\mathbf{y}^T$.

Show that $A = \lambda_1\mathbf{x}_1\mathbf{y}_1^T + \cdots + \lambda_n\mathbf{x}_n\mathbf{y}_n^T$, where $\mathbf{x}_i$ is the eigenvector associated with eigenvalue λ_i and $\mathbf{y}_i$ is the left eigenvector associated with eigenvalue λ_i.

Hint. What is A^T?

Approach:

- Recall the concept of left eigenvectors and how they relate to the transpose of a matrix.
- Consider the properties of diagonalizable matrices and their transpose.

Right and Left Eigenvectors

- **Right Eigenvectors:** For a matrix A, a right eigenvector v corresponds to an eigenvalue λ, such that $Av = \lambda v$. They are the standard eigenvectors that we often compute and use in diagonalization.
- **Left Eigenvectors:** The left eigenvectors y of a matrix A are the eigenvectors of A^T, satisfying $y^T A = \lambda y^T$ for the same eigenvalue λ. In the context of SVD, they are associated with the decomposition $A = U\Sigma V^T$, forming the columns of U.
- **Distinction:** While right eigenvectors are associated with the action of A on vectors, left eigenvectors relate to the action of A^T. In diagonalizable matrices, they provide a dual basis that makes up the invertible matrix P in the diagonalization $A = PDP^{-1}$, where P contains the right eigenvectors, and P^{-1} contains the left eigenvectors in its rows.

Calculating A^T

Starting with $A = X\Lambda X^{-1}$, compute A^T:

$$
\begin{aligned}
A^T &= \left(X\Lambda X^{-1}\right)^T \\
&= \left(X^{-1}\right)^T \Lambda^T X^T \\
&= \left(X^{-1}\right)^T \Lambda X^T
\end{aligned}
$$

- Note: Λ is diagonal, so $\Lambda^T = \Lambda$.
- This shows A^T in a form similar to diagonalization, but with $\left(X^{-1}\right)^T$ and X^T.

Eigenvectors of A^T

From $A^T = \left(X^{-1}\right)^T \Lambda X^T$, we identify eigenvectors:

- The columns of $\left(X^{-1}\right)^T$ are the eigenvectors of A^T.
- Equivalently, the rows of X^{-1} serve as the left eigenvectors of A.

This establishes a relationship between the diagonalization of A and the eigenvectors of A^T.

Decomposing A

$A = X\Lambda X^{-1}$ can be written as a sum of rank 1 matrices:

$$A = \sum_{i=1}^{n} \lambda_i x_i {y_i}^T$$

- Here, x_i are columns of X, and ${y_i}^T$ are rows of X^{-1}, corresponding to the left eigenvectors.
- λ_i are the eigenvalues.

This decomposition highlights the role of eigenvalues and (left) eigenvectors in forming A.