Singular Value Decomposition

Singular Value Decomposition (SVD) is a kind of generalized PCA (Principal Components Analysis) to select the top k principal components.

Compact/Reduced/Thin SVD: For any real matrix $\boldsymbol{A} \in \mathbb{R}^{n\times d}$ with rank r, there exist matrices:

$$
\begin{aligned}
\boldsymbol{U} &= \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r \end{bmatrix} \in \mathbb{R}^{n\times r}, \quad \boldsymbol{u}_i \in \mathbb{R}^n, \quad i = 1, 2, \dots, r \\
\boldsymbol{V} &= \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix} \in \mathbb{R}^{d\times r}, \quad \boldsymbol{v}_i \in \mathbb{R}^d, \quad i = 1, 2, \dots, r \\
\boldsymbol{\Sigma} &= \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \in \mathbb{R}^{r\times r}
\end{aligned}
$$

such that

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

where

- $\boldsymbol{U}$ and $\boldsymbol{V}$ have orthonormal columns, i.e.,

$$
\begin{aligned}
\boldsymbol{U}^\top\boldsymbol{U} &= \boldsymbol{I}_r \\
\boldsymbol{V}^\top\boldsymbol{V} &= \boldsymbol{I}_r
\end{aligned}
$$

  or equivalently, for $i, j = 1, 2, \dots, r$,

$$
\begin{aligned}
\boldsymbol{u}_i^\top\boldsymbol{u}_j &= \delta_{ij} \\
\boldsymbol{v}_i^\top\boldsymbol{v}_j &= \delta_{ij}
\end{aligned}
$$

- $\boldsymbol{\Sigma}$ is a diagonal matrix whose diagonal entries satisfy

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$$

Here,

- $\boldsymbol{u}_i$ is called the i-th left singular vector of $\boldsymbol{A}$.
- $\boldsymbol{v}_i$ is called the i-th right singular vector of $\boldsymbol{A}$.
- σ_i is called the i-th singular value of $\boldsymbol{A}$.

Connection between SVD and Eigenvalue Decomposition

- For any matrix $\boldsymbol{A}$,
  - The right singular vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_r\}$ are the eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$ (corresponding to its nonzero eigenvalues).
  - The left singular vectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r\}$ are the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^\top$ (corresponding to its nonzero eigenvalues)

- The singular values $\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ are the square roots of the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$ (equivalently, $\boldsymbol{A}\boldsymbol{A}^\top$):

$$\lambda_i = \sigma_i^2, \quad i = 1, 2, \ldots, r$$

where $r = \text{rank}(\boldsymbol{A})$.

- For a diagonalizable matrix $\boldsymbol{A}$,
  - $\boldsymbol{A}$ has n linearly independent eigenvectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ with eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, which satisfy:

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}$$

where,

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix},$$
$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Equivalently, the eigendecomposition of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$$

- For a real symmetric matrix $\boldsymbol{A}$,
  - $\boldsymbol{A}$ has n orthonormal eigenvectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ with eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, which satisfy:

$$\boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_n$$

  - Since real symmetric implies diagonalizable, the eigendecomposition of $\boldsymbol{A}$ becomes

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} \\ &= \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top \end{aligned}$$

- For a positive semidefinite (PSD) matrix $\boldsymbol{A}$,
  - Since $\boldsymbol{A}$ is PSD, all eigenvalues are nonnegative:

$$\lambda_i \geq 0$$

  - Since PSD implies real symmetric, $\boldsymbol{A}$ has the eigendecomposition

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$$

  - Thus, the eigendecomposition is also an SVD of $\boldsymbol{A}$,

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \\ &= \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top \end{aligned}$$

Here, the left and right singular vectors can be chosen to coincide:

$$\boldsymbol{U} = \boldsymbol{V}$$

and the singular values equal the eigenvalues,

$$\sigma_i = \lambda_i \geq 0, \quad \forall i$$

Geometric Meaning of SVD

A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times d}$ defines a linear transformation:

$$\boldsymbol{A} : \mathbb{R}^d \to \mathbb{R}^n$$

which maps any vector $\boldsymbol{x} \in \mathbb{R}^d$ to a new vector $\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^n$.

Since $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, this linear transformation can be decomposed into the three steps,

$$\boldsymbol{x} \xrightarrow{\text{Step 1}} \boldsymbol{V}^\top \boldsymbol{x} \xrightarrow{\text{Step 2}} \boldsymbol{\Sigma}\boldsymbol{V}^\top \boldsymbol{x} \xrightarrow{\text{Step 3}} \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}$$

These three steps are interpreted as follows:

- Step 1: Projection onto the right-singular subspace → the coordinate representation w.r.t. the right-singular-vector basis.
  - Let $\mathscr{V}$ denote the r-dimensional subspace spanned by the right singular vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_r\}$, i.e., $\mathscr{V} = \text{span}(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_r)$.

    The projection $\boldsymbol{x}$ onto $\mathscr{V}$ produces $\text{Proj}_{\mathscr{V}}(\boldsymbol{x})$, whose coordinate vector w.r.t. the basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_r\}$ is $\boldsymbol{V}^\top \boldsymbol{x}$.
    * $\boldsymbol{x}$ is the coordinate vector of $\boldsymbol{x}$ w.r.t. the standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_d\}$, where $\boldsymbol{e}_i \in \mathbb{R}^d$.

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{I}_d\, \boldsymbol{x} \\ &= \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \end{aligned}$$

    * $\boldsymbol{V}^\top \boldsymbol{x}$ is the coordinate vector of $\text{Proj}_{\mathscr{V}}(\boldsymbol{x})$ w.r.t. the basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_r\}$. Specifically, by orthogonal projection,

$$\begin{aligned} \text{Proj}_{\mathscr{V}}(\boldsymbol{x}) &= \sum_{i=1}^{r} \left(\boldsymbol{x}^\top \boldsymbol{v}_i\right) \boldsymbol{v}_i \\ &= \sum_{i=1}^{r} \left(\boldsymbol{v}_i^\top \boldsymbol{x}\right) \boldsymbol{v}_i \\ &= \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^\top \boldsymbol{x} \\ \boldsymbol{v}_2^\top \boldsymbol{x} \\ \vdots \\ \boldsymbol{v}_r^\top \boldsymbol{x} \end{bmatrix} \end{aligned}$$

Here,

$$\begin{aligned}\begin{bmatrix} \boldsymbol{v}_1^\top \boldsymbol{x} \\ \boldsymbol{v}_2^\top \boldsymbol{x} \\ \vdots \\ \boldsymbol{v}_r^\top \boldsymbol{x} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{v}_1^\top \\ \boldsymbol{v}_2^\top \\ \vdots \\ \boldsymbol{v}_r^\top \end{bmatrix} \boldsymbol{x} \\ &= \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r \end{bmatrix}^\top \boldsymbol{x} \\ &= \boldsymbol{V}^\top \boldsymbol{x}\end{aligned}$$

- When $r < d$, this yields an r-dimensional coordinate representation of $\text{Proj}_{\mathscr{V}}(\boldsymbol{x})$, reducing the representation dimension from d to r.

- Step 2: Scaling the coordinate vector by singular values to obtain the scaled coordinate vector.
  - $\boldsymbol{\Sigma}\boldsymbol{V}^\top \boldsymbol{x}$ scales each component $\boldsymbol{v}_i^\top \boldsymbol{x}$ of the coordinate vector $\boldsymbol{V}^\top \boldsymbol{x}$ independently by the corresponding singular value σ_i, i.e.,

$$\boldsymbol{\Sigma}\boldsymbol{V}^\top \boldsymbol{x} = \begin{bmatrix} \sigma_1 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^\top \boldsymbol{x} \\ \boldsymbol{v}_2^\top \boldsymbol{x} \\ \vdots \\ \boldsymbol{v}_r^\top \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \sigma_1 \boldsymbol{v}_1^\top \boldsymbol{x} \\ \sigma_2 \boldsymbol{v}_2^\top \boldsymbol{x} \\ \vdots \\ \sigma_r \boldsymbol{v}_r^\top \boldsymbol{x} \end{bmatrix}$$

- Step 3: Reconstruction of the final vector from the scaled coordinate vector w.r.t. the left-singular-vector basis.
  - Let $\mathscr{U}$ denote the r-dimensional subspace spanned by the left singular vectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$, i.e., $\mathscr{U} = \text{span}(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r)$. $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \boldsymbol{x}$ reconstructs the final vector $\boldsymbol{A}\boldsymbol{x} \in \mathscr{U}$ as a linear combination of the basis vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r$, with the scaled coordinate vector $\boldsymbol{\Sigma}\boldsymbol{V}^\top \boldsymbol{x}$ as the coefficient vector, i.e.,

$$\begin{aligned}\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \boldsymbol{x} \\ &= \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 \boldsymbol{v}_1^\top \boldsymbol{x} \\ \sigma_2 \boldsymbol{v}_2^\top \boldsymbol{x} \\ \vdots \\ \sigma_r \boldsymbol{v}_r^\top \boldsymbol{x} \end{bmatrix}\end{aligned}$$

Geometric Meaning Eigenvalues and Eigenvectors

- A square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ defines a linear transformation:

$$\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^n$$

  which maps any vector $\boldsymbol{x} \in \mathbb{R}^n$ to a new vector $\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^n$.

- An eigenvector is a nonzero vector whose direction is preserved under a linear transformation. An eigenvalue is the scalar corresponding to an eigenvector that determines how the eigenvector is transformed under the linear transformation: its magnitude specifies the scaling factor, while its sign determines whether the orientation is preserved or reversed.

Specifically, let λ be an eigenvalue of $\boldsymbol{A}$ with corresponding eigenvector $\boldsymbol{v}$. Then

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v},$$

where

- If $\lambda > 0$, the transformed vector $\lambda\boldsymbol{v}$ has the same direction as the eigenvector $\boldsymbol{v}$.
- If $\lambda < 0$, the transformed vector has the opposite direction to the eigenvector.
- If $\lambda = 0$, the transformed vector is the zero vector.

• For any vector parallel to an eigenvector, a linear transformation maps it to another vector that is still parallel to the eigenvector and scaled by the corresponding eigenvalue.

Specifically, let $\boldsymbol{v}$ be an eigenvector of $\boldsymbol{A}$ with corresponding eigenvalue λ.

Any vector $\boldsymbol{x} \in \mathbb{R}^n$ parallel to $\boldsymbol{v}$ can be written as

$$\boldsymbol{x} = k\boldsymbol{v} \in \mathbb{R}^n$$

Then,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}(k\boldsymbol{v}) = k(\boldsymbol{A}\boldsymbol{v}) = k(\lambda\boldsymbol{v}) = \lambda(k\boldsymbol{v}) = \lambda\boldsymbol{x}$$

• For any vector, the linear transformation (suppose the transformation matrix has n linearly independent eigenvectors with n corresponding eigenvalues) maps it to another vector in three steps:

- Decomposition: Express the vector as a linear combination of the eigenvectors.
- Transformation: Scale each eigenvector component by its corresponding eigenvalue.
- Summation: Sum the scaled components to obtain the final vector.

Specifically, let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be the n eigenvectors of $\boldsymbol{A}$, and $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ be the corresponding eigenvalues.

Any vector $\boldsymbol{x} \in \mathbb{R}^n$ can be written as a linear combination of these eigenvectors,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\sum_{i=1}^{n} k_i\,\boldsymbol{v}_i$$

Then,

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{A}\sum_{i=1}^{n} k_i\,\boldsymbol{v}_i \\
&= \sum_{i=1}^{n} k_i\left(\boldsymbol{A}\boldsymbol{v}_i\right) \\
&= \sum_{i=1}^{n} k_i\left(\lambda_i\,\boldsymbol{v}_i\right) \\
&= \sum_{i=1}^{n} \lambda_i\left(k_i\,\boldsymbol{v}_i\right)
\end{aligned}$$