Mathematics for Machine Learning

Vazgen Mikayelyan

August 4, 2020

FAST DISCOVERING THE FUTURE

Eigendecomposition and Diagonalization

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ *is* **diagonalizable** *if it is similar to a diagonal matrix, i.e. there exists a matrix* $P \in \mathbb{R}^{n\times n}$ *so that*

$$D = P^{-1}AP,$$

(equivalently $AP = PD$ *or* $A = PDP^{-1}$*) where* $D = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}$

Eigendecomposition and Diagonalization

Proposition

If $D \in \mathbb{R}^{n \times n}$ *is a diagonal matrix with diagonal entries* $\lambda_1, \ldots, \lambda_n$*, then*

$$AP = PD$$

for invertible matrix P *if and only if* $\lambda_1, \ldots, \lambda_n$ *are the eigenvalues of* A *and the* $\mathbf{p}_i$ *are the corresponding eigenvectors of* A*, where* $P = [\mathbf{p}_1 \ldots \mathbf{p}_n]$*.*

Theorem

A symmetric matrix $S = S^T \in \mathbb{R}^{n \times n}$ *can be diagonalized into*

$$S = PDP^T$$

where P *is matrix of* n *orthogonal eigenvectors,i.e.* $P^T = P^{-1}$,
and D *is a diagonal matrix of its* n *eigenvalues.*

Theorem

A symmetric matrix $S = S^T \in \mathbb{R}^{n \times n}$ *can be diagonalized into*

$$S = PDP^T$$

where P *is matrix of* n *orthogonal eigenvectors,i.e.* $P^T = P^{-1}$, *and* D *is a diagonal matrix of its* n *eigenvalues.*

Example

Compute the eigendecomposition of a (symmetric) matrix $A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

Theorem

A symmetric matrix $S = S^T \in \mathbb{R}^{n\times n}$ can be diagonalized into

$$S = PDP^T$$

where P is matrix of n orthogonal eigenvectors,i.e. $P^T = P^{-1}$, and D is a diagonal matrix of its n eigenvalues.

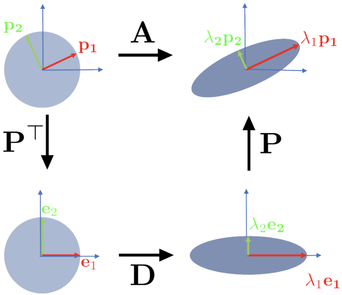
Example

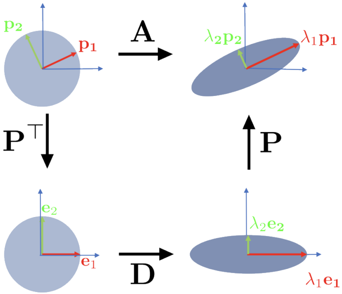
Compute the eigendecomposition of a (symmetric) matrix $A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

Proposition

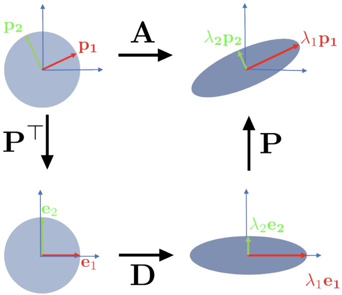
If $A = PDP^{-1}$, then $A^k = PD^kP^{-1}$ for any $k \in N$, and

$$D^k = \begin{bmatrix} \lambda_1^k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^k \end{bmatrix}$$

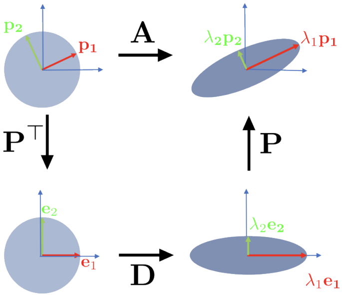

$\mathbf{p_2}$
$\mathbf{p_1}$
$\mathbf{A}$
$\lambda_2\mathbf{p_2}$
$\lambda_1\mathbf{p_1}$
$\mathbf{P}^\top$
$\mathbf{P}$
$\mathbf{e}_2$
$\mathbf{e}_1$
$\mathbf{D}$
$\lambda_2\mathbf{e_2}$
$\lambda_1\mathbf{e_1}$



- Top-left to bottom-left: P^{-1} performs a basis change (here drawn in R^2 and depicted as a rotation-like operation), mapping the eigenvectors into the standard basis.



- Top-left to bottom-left: P^{-1} performs a basis change (here drawn in R^2 and depicted as a rotation-like operation), mapping the eigenvectors into the standard basis.
- Bottom-left to bottom-right: D performs a scaling along the remapped orthogonal eigenvectors, depicted here by a circle being stretched to an ellipse.



- Top-left to bottom-left: P^{-1} performs a basis change (here drawn in R^2 and depicted as a rotation-like operation), mapping the eigenvectors into the standard basis.
- Bottom-left to bottom-right: D performs a scaling along the remapped orthogonal eigenvectors, depicted here by a circle being stretched to an ellipse.
- Bottom-right to top-right: P undoes the basis change (depicted as a reverse rotation) and restores the original coordinate frame.

Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD)

Theorem

Let $A \in \mathbb{R}^{m \times n}$ *be a rectangular matrix of rank* r, *with* $r \in [0, \min(m,n)]$. *The Singular Value Decomposition or SVD of* A *is a decomposition of* A *of the form*

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

where $U \in \mathbb{R}^{m \times m}$ *is an orthogonal matrix of column vectors* $\mathbf{u}_i$, *and* $V \in \mathbb{R}^{n \times n}$ *is an orthogonal matrix of column vectors* $\mathbf{v}_j$ *and* Σ *is an* $m \times n$ *matrix with* $\Sigma_{ii} = \sigma_i > 0$ *and* $\Sigma_{ij} = 0,\ i \neq j$. *The SVD is always possible for any matrix* A.

Singular Value Decomposition (SVD)

Theorem

Let $A \in \mathbb{R}^{m\times n}$ be a rectangular matrix of rank r, with $r \in [0, \min(m,n)]$. The Singular Value Decomposition or SVD of A is a decomposition of A of the form

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

where $U \in \mathbb{R}^{m\times m}$ is an orthogonal matrix of column vectors $\mathbf{u}_i$, and $V \in \mathbb{R}^{n\times n}$ is an orthogonal matrix of column vectors $\mathbf{v}_j$ and Σ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i > 0$ and $\Sigma_{ij} = 0,\ i \neq j$. The SVD is always possible for any matrix A.

The σ_i are called the singular values, and by convention the singular values are ordered, i.e., $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$.

Singular Value Decomposition (SVD)

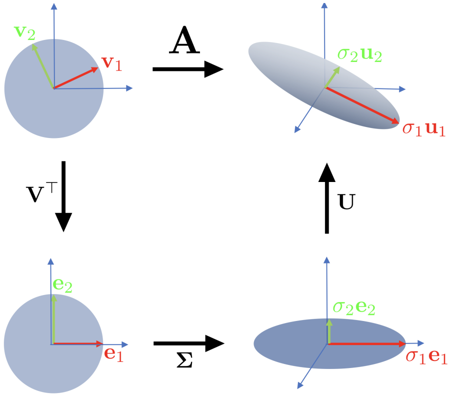
Theorem

Let $A \in \mathbb{R}^{m \times n}$ be a rectangular matrix of rank r, with $r \in [0, \min(m, n)]$. The Singular Value Decomposition or SVD of A is a decomposition of A of the form

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

where $U \in \mathbb{R}^{m \times m}$ is an orthogonal matrix of column vectors $\mathbf{u}_i$, and $V \in \mathbb{R}^{n \times n}$ is an orthogonal matrix of column vectors $\mathbf{v}_j$ and Σ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i > 0$ and $\Sigma_{ij} = 0, \ i \neq j$. The SVD is always possible for any matrix A.

The σ_i are called the singular values, and by convention the singular values are ordered, i.e., $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$. $\mathbf{u}_i$ are called the left-singular vectors and $\mathbf{v}_j$ are called the right-singular vectors.


$\mathbf{v}_2$
$\mathbf{v}_1$
$\mathbf{A}$
$\sigma_2 \mathbf{u}_2$
$\sigma_1 \mathbf{u}_1$
$\mathbf{V}^\top$
$\mathbf{U}$
$\mathbf{e}_2$
$\mathbf{e}_1$
$\boldsymbol{\Sigma}$
$\sigma_2 \mathbf{e}_2$
$\sigma_1 \mathbf{e}_1$

Σ has a diagonal submatrix that contains the singular values and needs additional zero vectors that increase the dimension.

if $m > n$

$$\Sigma = \begin{bmatrix} \sigma_1 & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \sigma_n \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

if $m < n$

$$\Sigma = \begin{bmatrix} \sigma_1 & \dots & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 & \dots & 0 \\ 0 & \dots & \sigma_n & 0 & \dots & 0 \end{bmatrix}$$

The singular value matrix Σ must be of the same size as A

Example

$$A = \begin{pmatrix} -1 & \frac{1}{\sqrt{2}} & 1 \\ -1 & -\frac{1}{\sqrt{2}} & 1 \end{pmatrix} = U\Sigma V^T$$

$$= \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & -1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{pmatrix}$$

Example

$$A = \begin{pmatrix} -1 & \frac{1}{\sqrt{2}} & 1 \\ -1 & -\frac{1}{\sqrt{2}} & 1 \end{pmatrix} = U\Sigma V^T$$

$$= \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & -1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \end{pmatrix}$$

Example

$$A = \begin{pmatrix} -\frac{3\sqrt{3}}{4} & -\frac{3}{4} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{9}{4} & -\frac{3\sqrt{3}}{4} \end{pmatrix} = U\Sigma V^T$$

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

Remark

The eigenvalue decomposition of a symmetric matrix

$$S = S^T = PDP^T$$

is a special case of the SVD

$$S = U\Sigma V^T,$$

where

Remark

The eigenvalue decomposition of a symmetric matrix

$$S = S^T = PDP^T$$

is a special case of the SVD

$$S = U\Sigma V^T,$$

where $U = V = P$ *and* $\Sigma = D$.

Let $r = \mathrm{rank}(A)$.

Remark

- *The columns of* U *(m by m) are eigenvectors of* AA^T,
- *the columns of* V *(n by n) are eigenvectors of* A^TA.
- *The* r *singular values on the diagonal of* Σ *(m by n) are the square roots of the nonzero eigenvalues of both* AA^T *and* A^TA.

Let $r = \operatorname{rank}(A)$.

Remark

- *The columns of U (m by m) are eigenvectors of AA^T,*
- *the columns of V (n by n) are eigenvectors of $A^T A$.*
- *The r singular values on the diagonal of Σ (m by n) are the square roots of the nonzero eigenvalues of both AA^T and $A^T A$.*

Remark

U and V give orthonormal bases for all four fundamental subspaces:

- *first r columns of U: column space of A*
- *last $m - r$ columns of U: nullspace of A^T*
- *first r columns of V: row space of A*
- *last $n - r$ columns of V: nullspace of A*

Remark

When A multiplies a column $\mathbf{v}_i$ of V, it produces σ_i times a column of U.

$$AV = U\Sigma \Leftrightarrow A\mathbf{v}_i = \sigma_i \mathbf{u}_i$$

V 1. How to compute the SVD in **the case rank**$(A) = m \leq n$ **?**

Step 1: Compute the symmetrized matrix $A^T A$ (recall $A \in \mathbb{R}^{m\times n}$).
Step 2: Compute the eigenvalue decomposition of $A^T A = PDP^T$. From here we obtain

$$V = P, \text{ and } \Sigma^T \Sigma = D,$$

The eigenvalues of $A^T A$ are the squared singular values of Σ.
Step 3. Compute U using the formula

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i, \quad i = 1, \ldots, m(m = \text{rank}(A))$$

V 1. How to compute the SVD in **the case rank**$(A) = m \leq n$ **?**

Step 1: Compute the symmetrized matrix $A^T A$ (recall $A \in \mathbb{R}^{m \times n}$).
Step 2: Compute the eigenvalue decomposition of $A^T A = PDP^T$. From here we obtain
$$V = P, \text{ and } \Sigma^T \Sigma = D,$$

The eigenvalues of $A^T A$ are the squared singular values of Σ.
Step 3. Compute U using the formula

$$\mathbf{u}_i = \frac{1}{\sigma_i} A \mathbf{v}_i, \quad i = 1, \ldots, m (m = \text{rank}(A))$$

V 2. How to compute the SVD in **the case rank**$(A) = n \leq m$**?**

Step 1: Compute the symmetrized matrix AA^T.
Step 2: Compute the eigenvalue decomposition of $AA^T = QD_1Q^T$. From here we obtain $U = Q$, and $\Sigma\Sigma^T = D_1$,
Step 3. Compute V using the formula

$$\mathbf{v}_i = \frac{1}{\sigma_i} A^T \mathbf{u}_i, \quad i = 1, \ldots, n (n = \text{rank}(A))$$

V 3. How to compute the SVD in **general case** (**rank**$(A) = r \leq \min(m,n)$)?

Step 1: Compute the symmetrized matrix AA^T.
Step 2: Compute the eigenvalue decomposition of $AA^T = QD_1Q^T$. From here we obtain $U = Q$, and $\Sigma\Sigma^T = D_1$,
Step 3. Compute V using the formula

$$\mathbf{v}_i = \frac{1}{\sigma_i} A^T \mathbf{u}_i, \quad \text{for } i = 1, \ldots, r$$

and choose $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n$ so that they form an orthonormal basis of the nullspace of A

Example

Find the SVD of the matrix $A = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

Example

Find the SVD of the matrix $A = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

Answer

$$A = U\Sigma V^T =$$
$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Example

Find the SVD of the matrix $A = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

Answer

$$A = U\Sigma V^T =$$
$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Remark

If we were asked to find the SVD of the transpose of the initial matrix, i.e. $A^T = \begin{pmatrix} 0 & 1 \\ 2 & 0 \\ 0 & 0 \end{pmatrix}$*, then we would use the 2nd version of "How to compute the SVD?", as in that case rank*$(A) = 2 =$*number of columns. So we would first find the* **three** *left singular vectors (new* $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$*), then using those by* $\mathbf{v}_i = \frac{1}{\sigma_i} A\mathbf{u}_i$*, we would find the* **two** *right singular vectors (new* $\mathbf{v}_1, \mathbf{v}_2$*).*

Example

Find the SVD of the matrix $A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix}$

Example

Find the SVD of the matrix $A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{pmatrix}$

Answer

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

$$\begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{\sqrt{2}}{6} & \frac{2}{3} \\ -\frac{1}{\sqrt{2}} & \frac{\sqrt{2}}{6} & -\frac{2}{3} \\ 0 & -\frac{2\sqrt{2}}{3} & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

Mathematical Analysis

Limit of a Sequence

Limit of a Sequence

Definition

We call $x \in \mathbb{R}$ *the limit of the sequence* $\{x_n\}_{n=1}^{\infty}$ *if the following condition holds: for each real number* $\varepsilon > 0$*, there exists a natural number* n_0 *such that, for every natural number* $n \geq n_0$*, we have* $|x_n - x| < \varepsilon$*.*

Limit of a Sequence

Definition

We call $x \in \mathbb{R}$ *the limit of the sequence* $\{x_n\}_{n=1}^{\infty}$ *if the following condition holds: for each real number* $\varepsilon > 0$*, there exists a natural number* n_0 *such that, for every natural number* $n \geq n_0$*, we have* $|x_n - x| < \varepsilon$*.*

We will write $\lim\limits_{n \to \infty} x_n = x$ or $x_n \to x$.

Limit of a Sequence

Definition

We call $x \in \mathbb{R}$ *the limit of the sequence* $\{x_n\}_{n=1}^{\infty}$ *if the following condition holds: for each real number* $\varepsilon > 0$*, there exists a natural number* n_0 *such that, for every natural number* $n \geq n_0$*, we have* $|x_n - x| < \varepsilon$*.*

We will write $\lim\limits_{n \to \infty} x_n = x$ or $x_n \to x$.

Definition

We will say than $\{x_n\}_{n=1}^{\infty}$ *tends to infinity if the following condition holds: for each real number* E*, there exists a natural number* n_0 *such that, for every natural number* $n \geq n_0$*, we have* $x_n > E$ *.*

Limit of a Sequence

Definition

We call $x \in \mathbb{R}$ the limit of the sequence $\{x_n\}_{n=1}^{\infty}$ if the following condition holds: for each real number $\varepsilon > 0$, there exists a natural number n_0 such that, for every natural number $n \geq n_0$, we have $|x_n - x| < \varepsilon$.

We will write $\lim\limits_{n\to\infty} x_n = x$ or $x_n \to x$.

Definition

We will say than $\{x_n\}_{n=1}^{\infty}$ tends to infinity if the following condition holds: for each real number E, there exists a natural number n_0 such that, for every natural number $n \geq n_0$, we have $x_n > E$.

We will write $\lim\limits_{n\to\infty} x_n = +\infty$ or $x_n \to +\infty$.