Machine Learning & Pattern Recognition

SONG Xuemeng

sxmustc@gmail.com

http://xuemeng.bitcron.com/

Review of Linear Algebra

- Vectors
- Products and norms
- Linear dependence and independence
- Vector spaces and basis
- Matrices
- Linear transformations
- Eigenvalues and eigenvectors

Vectors

- An d-dimensional column vector and its transpose (row vector) are represented as

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \qquad \boldsymbol{x}^{\mathrm{T}} = [x_1 \; x_2 \cdots \; x_d]$$

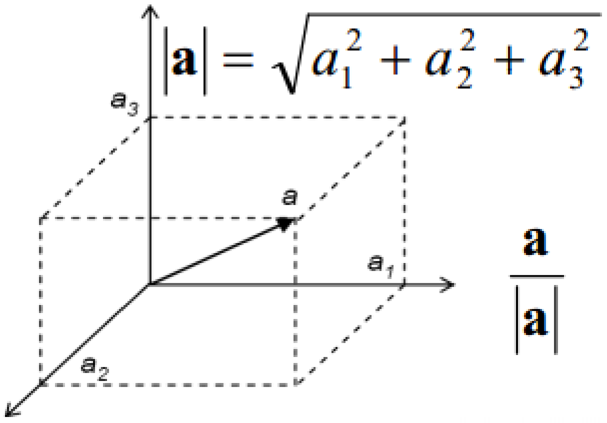
- The inner product (dot product) of two vectors:

$$< \boldsymbol{x}, \boldsymbol{y} >= \mathbf{x}^T \boldsymbol{y} = \boldsymbol{y}^T \boldsymbol{x} = \sum_{k=1}^{d} x_k y_k$$

Vectors

- Euclidean norm or length

$$|\boldsymbol{x}| = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \left[\sum_{k=1}^{d} x_k x_k\right]^{1/2}$$

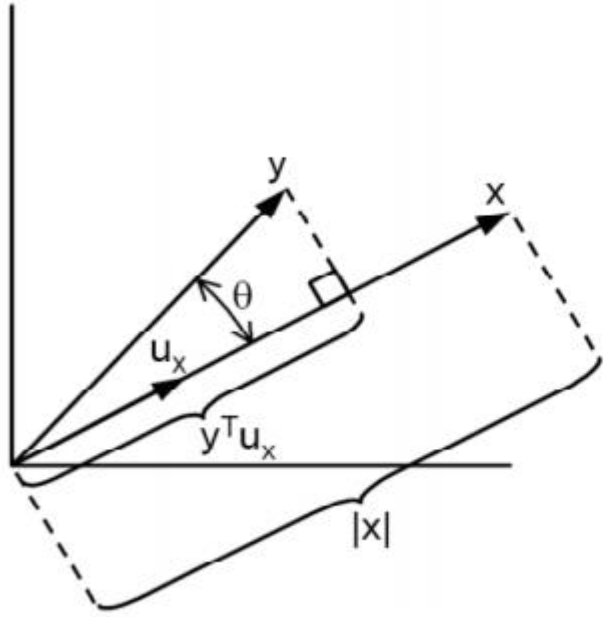


- Normalized (unit) vector

$$|\boldsymbol{x}| = 1$$

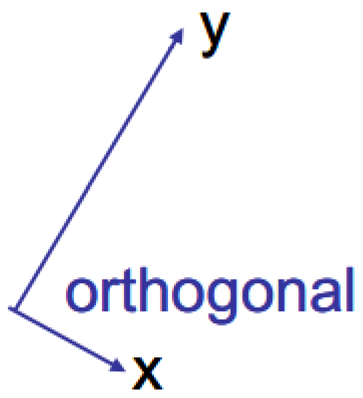
- Angle between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$

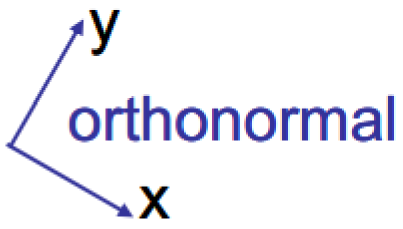
$$\cos\theta = \frac{<\boldsymbol{x},\boldsymbol{y}>}{|\boldsymbol{x}|\cdot|\boldsymbol{y}|}$$



Vectors

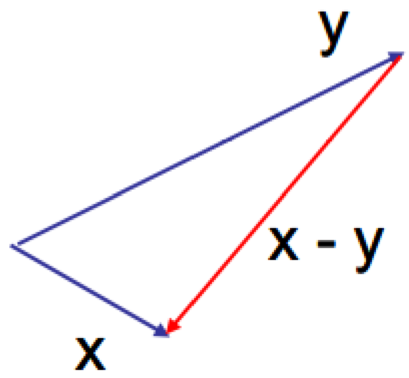
- Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are
  - *Orthogonal* if $\cos\theta = 0$ or $< \boldsymbol{x}, \boldsymbol{y} >= \mathbf{0}$
  - *Orthonormal* if they are orthogonal and $|\boldsymbol{x}| = |\boldsymbol{y}| = 1$





- Euclidean distance between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$

$$|\boldsymbol{x} - \boldsymbol{y}| = \sqrt{\sum_{k=1}^{d} (x_k - y_k)^2}$$



Vector Norms

- A function $||\cdot||: \mathbb{R}^d \rightarrow \mathbb{R}$ is called a vector norm if it has the following properties:

1. $||\mathbf{x}|| \geq 0$ for any vector $\mathbf{x} \in \mathbb{R}^d$, and $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
2. $||\alpha\mathbf{x}|| = \alpha||\mathbf{x}||$ for any vector $\mathbf{x} \in \mathbb{R}^d$ and any scalar $\alpha \in \mathbb{R}$
3. $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$

- The most commonly used vector norms is the family of *p- norms*, or *l_p-norms*, defined by

$$||\mathbf{x}||_p = (\sum_{i=1}^{d} |x_i|^p)^{1/p} \qquad \mathbf{x} \in \mathbb{R}^d$$

Vector Norms

- The most commonly used vector norms belong to the family of p-$norms$, or l_p-$norms$, defined by

$$\|\mathbf{x}\|_p = (\sum_{i=1}^{d} |x_i|^p)^{1/p} \qquad \mathbf{x} \in \mathbb{R}^d$$

- The following l_p-$norms$ are of particular interest:
- $p = 1$: the l_1-$norm$

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_d|$$

Vector Norms

- The most commonly used vector norms belong to the family of p-$norms$, or l_p-$norms$, defined by

$$\|\mathbf{x}\|_p = (\sum_{i=1}^{d} |x_i|^p)^{1/p} \qquad \mathbf{x} \in \mathbb{R}^d$$

- The following l_p-$norms$ are of particular interest:
- $p = 1$: the l_1-$norm$

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_d|$$

- $p = 2$: the l_2-$norm$ or *Euclidean norm*

$$\|\mathbf{x}\|_2 = \sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_d}^2}$$

Vector Norms

- The most commonly used vector norms belong to the family of p-$norms$, or l_p-$norms$, defined by

$$\|\mathbf{x}\|_p = (\sum_{i=1}^{d} |x_i|^p)^{1/p} \qquad \mathbf{x} \in \mathbb{R}^d$$

- The following l_p-$norms$ are of particular interest:
- $p = 1$: the l_1-$norm$

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_d|$$

- $p = 2$: the l_2-$norm$ or *Euclidean norm*

$$\|\mathbf{x}\|_2 = \sqrt{{x_1}^2 + {x_2}^2 + \cdots + {x_d}^2}$$

- $p = \infty$: the l_∞-$norm$

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le d} |x_i|$$

Linear Dependence and Independence

- Vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are linearly **dependent** if there exists a set of coefficients $a_1, a_2, \ldots, a_n$ (at least one $a_i \neq 0$) such that

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n = 0$$

- Vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are linearly **independent** if

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n = 0 \Rightarrow a_k = 0, \quad \forall k$$

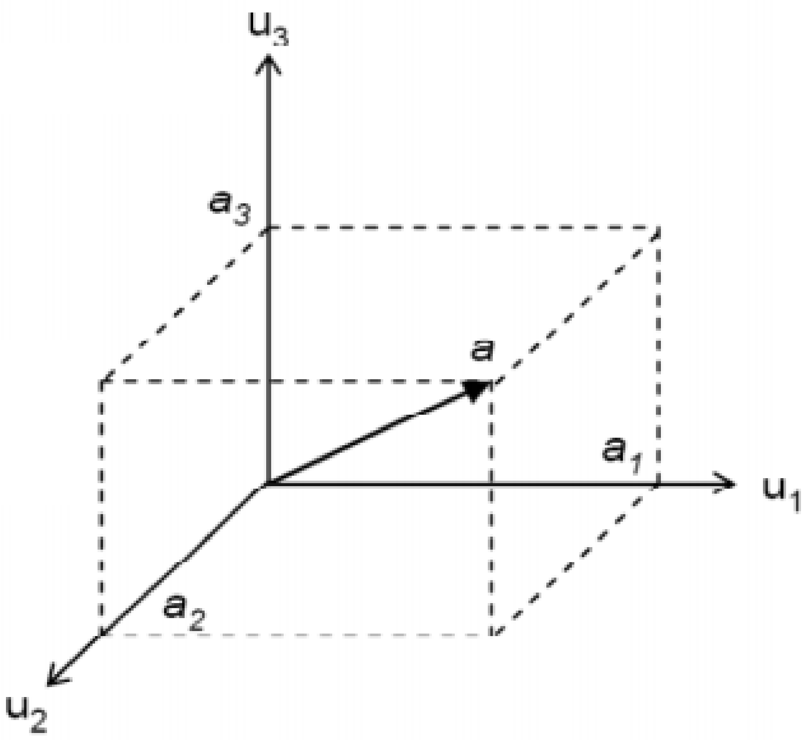
- **Vector Space:**
  - The n-dimensional space, in which all the n-dimensional vectors reside.

Vector Spaces and Basis

- **Basis**
  - A set of vectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n\}$ are called a ***basis*** for a vector space if any vector $\boldsymbol{x}$ can be written as a linear combination of $\{\boldsymbol{u}_i\}$

$$\boldsymbol{x} = a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2 + \cdots + a_n\boldsymbol{u}_n$$

  - $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$ are **independent** implies they form a basis and vice versa.
  - A basis $\{\boldsymbol{u}_i\}$ is **orthonormal** if
    - Basis vectors are pairwise orthogonal
    - Have unit length, i.e., $|\boldsymbol{u}_i| = 1.$



Orthonormal basis

Matrices

- An n by d matrix A and its **transpose** A^{T}

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1d} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & & a_{nd} \end{bmatrix}_{n \times d} \qquad A^{T} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ a_{13} & a_{23} & \cdots & a_{n3} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1d} & a_{2d} & & a_{nd} \end{bmatrix}_{d \times n}$$

- Product of two matrices:

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1d} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2d} \\ \vdots & \vdots & \vdots & \ddots & \\ a_{m1} & a_{m2} & a_{m3} & & a_{md} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ b_{31} & b_{32} & \cdots & b_{3n} \\ \vdots & \vdots & \ddots & \\ b_{d1} & b_{d2} & & b_{dn} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2n} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \\ c_{m1} & c_{m2} & c_{m3} & & c_{mn} \end{bmatrix}$$

$$\boxed{c_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj}}$$

Matrices

- $(\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T$

- **Identity matrix:** $\boldsymbol{I}$

$$\boldsymbol{IA} = \boldsymbol{AI} = \boldsymbol{A}$$

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

- **Symmetric:** $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 8 & 6 \\ 2 & 3 & 4 & 9 \\ 8 & 4 & 5 & 6 \\ 6 & 9 & 6 & 7 \end{bmatrix}$$

Matrices

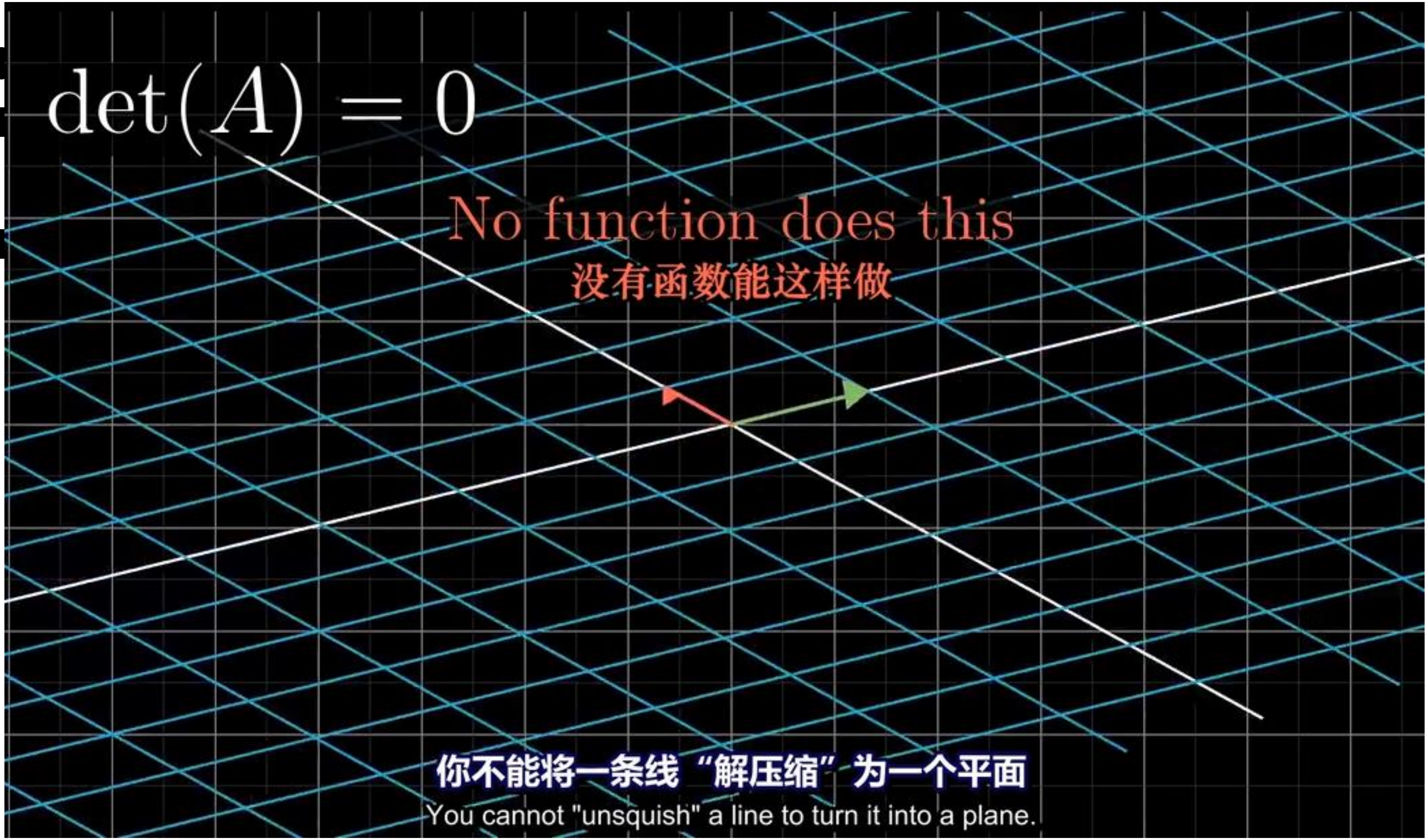
- **Inverse**
  - The inverse of a square matrix $\boldsymbol{A}$ is $\boldsymbol{A}^{-1}$

$$\boldsymbol{A}A^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

  - The inverse $\boldsymbol{A}^{-1}$ exists if and only if A is non-singular

$$|\boldsymbol{A}| \neq 0$$

Matrices



$\det(A) = 0$
No function does this
没有函数能这样做
你不能将一条线“解压缩”为一个平面
You cannot "unsquish" a line to turn it into a plane.

Matrices

- **Inverse**
  - The inverse of a square matrix $\boldsymbol{A}$ is $\boldsymbol{A}^{-1}$

$$\boldsymbol{A}A^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

  - The inverse $\boldsymbol{A}^{-1}$ exists if and only if A is non-singular

$$|\boldsymbol{A}| \neq 0$$

- $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$ $\quad\quad$ $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

Matrices

- **Inverse**
  - The inverse of a square matrix $\boldsymbol{A}$ is $\boldsymbol{A}^{-1}$

$$\boldsymbol{A}A^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$$

  - The inverse $\boldsymbol{A}^{-1}$ exists if and only if A is non-singular

$$|\boldsymbol{A}| \neq 0$$

- $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$ $\qquad$ $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

- **Pseudo-inverse**

$$\boldsymbol{A}^{\dagger} = [\boldsymbol{A}^T\boldsymbol{A}]^{-1}\boldsymbol{A}^T \text{ with } \boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{I}$$

  - Assuming $\boldsymbol{A}^T\boldsymbol{A}$ is non-singular
  - Used whenever $\boldsymbol{A}^{-1}$ does not exists, i.e., $\boldsymbol{A}$ is not square or $\boldsymbol{A}$ is singular.

Matrices

- For a **square** matrix $\boldsymbol{A}$
  - Positive definite:
    if
    $$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0 \text{ for all } \boldsymbol{x} \neq \boldsymbol{0}$$
  - Semi-positive definite:
    if
    $$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq 0 \text{ for all } \boldsymbol{x} \neq \boldsymbol{0}$$

Matrices

- For a **square** matrix $\boldsymbol{A}$
  Trace: sum of diagonal elements

$$tr(\boldsymbol{A}) = \sum_{k=1}^{d} a_{kk}$$

  - $tr(\boldsymbol{A}) = tr(\boldsymbol{A}^T)$
  - $tr(\alpha\boldsymbol{A} + \beta\boldsymbol{B}) = \alpha tr(\boldsymbol{A}) + \beta tr(\boldsymbol{B})$ (Linearity)
  - $tr(\boldsymbol{ABC}) = tr(\boldsymbol{BCA}) = tr(\boldsymbol{CAB})$ (Cyclic property)

Matrices

- For a **square** matrix $\boldsymbol{A}$
  Trace: sum of diagonal elements

$$tr(\boldsymbol{A}) = \sum_{k=1}^{d} a_{kk}$$

- $tr(\boldsymbol{A}) = tr(\boldsymbol{A}^T)$
- $tr(\alpha\boldsymbol{A} + \beta\boldsymbol{B}) = \alpha tr(\boldsymbol{A}) + \beta tr(\boldsymbol{B})$ (Linearity)
- $tr(\boldsymbol{ABC}) = tr(\boldsymbol{BCA}) = tr(\boldsymbol{CAB})$ (Cyclic property)

How about? $tr(\boldsymbol{AB})\ ?\ tr(\boldsymbol{BA})$ $tr(\boldsymbol{ABC})\ ?\ tr(\boldsymbol{ACB})$

Matrices

- For a **square** matrix $\boldsymbol{A}$
  Trace: sum of diagonal elements

$$tr(\boldsymbol{A}) = \sum_{k=1}^{d} a_{kk}$$

- $tr(\boldsymbol{A}) = tr(\boldsymbol{A}^T)$
- $tr(\alpha\boldsymbol{A} + \beta\boldsymbol{B}) = \alpha tr(\boldsymbol{A}) + \beta tr(\boldsymbol{B})$ (Linearity)
- $tr(\boldsymbol{ABC}) = tr(\boldsymbol{BCA}) = tr(\boldsymbol{CAB})$ (Cyclic property)

How about? $tr(\boldsymbol{AB})\ ?\ tr(\boldsymbol{BA})$ $\quad tr(\boldsymbol{ABC})\ ?\ tr(\boldsymbol{ACB})$

$=$ $\quad\neq$

Matrix Norms

- The Entrywise matrix norms are of particular interest:
- Treat an $m \times n$ matrix as a vector of size mn, and use one of the familiar vector norms.

$$\|\mathbf{A}\|_p = \|\mathrm{vec}(\mathbf{A})\|_p = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|^p\right)^{1/p}, \mathbf{A} \in \mathbb{R}^{m\times n}$$

Matrix Norms

- The Entrywise matrix norms are of particular interest:
- Treat an $m \times n$ matrix as a vector of size mn, and use one of the familiar vector norms.

$$\|\mathbf{A}\|_p = \|\mathrm{vec}(\mathbf{A})\|_p = (\sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|^p)^{1/p} \ , \mathbf{A} \in \mathbb{R}^{m\times n}$$

- $p = 2$: the Frobenius norm

$$\|\boldsymbol{A}\|_F = (tr(\boldsymbol{A}^T\boldsymbol{A}))^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$$

Matrix Norms

- The Entrywise matrix norms are of particular interest:
- Treat an $m \times n$ matrix as a vector of size mn, and use one of the familiar vector norms.

$$\|\mathbf{A}\|_p = \|\mathrm{vec}(\mathbf{A})\|_p = (\sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|^p)^{1/p} \ , \mathbf{A} \in \mathbb{R}^{m\times n}$$

- $p = 2$: the Frobenius norm

$$\|\boldsymbol{A}\|_F = (tr(\boldsymbol{A}^T\boldsymbol{A}))^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$$

- $l_{2,1}$ norm

$\boldsymbol{a}_i$: the i-th rows of matrix $\mathbf{A}$

$$\|\boldsymbol{A}\|_{2,1} = \sum_{i=1}^{m}\|\boldsymbol{a}_i\|_2 = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$$

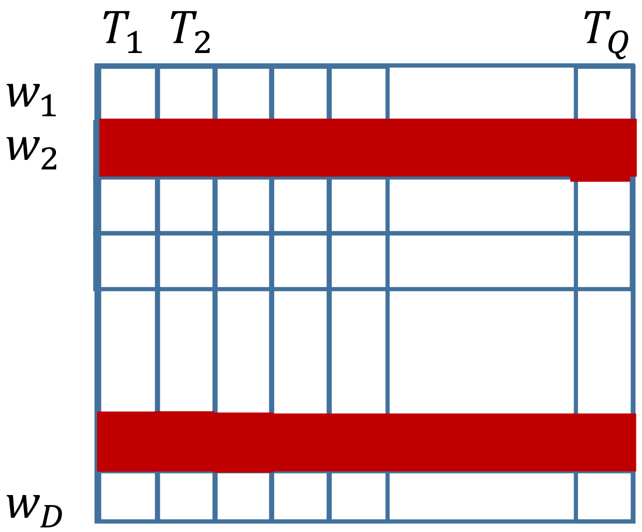
Matrix Norms

- The Entrywise matrix norms are of particular interest:
- Treat an $m \times n$ matrix as a vector of size mn, and use one of the familiar vector norms.

$$\|\mathbf{A}\|_p = \|\mathrm{vec}(\mathbf{A})\|_p = (\sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|^p)^{1/p} \;, \mathbf{A} \in \mathbb{R}^{m\times n}$$

- $p = 2$: the Frobenius norm

$$\|\boldsymbol{A}\|_F = (tr(\boldsymbol{A}^T\boldsymbol{A}))^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$$



- $l_{2,1}$ norm

$\boldsymbol{a}_i$: the i-th rows of matrix $\mathbf{A}$

$$\|\boldsymbol{A}\|_{2,1} = \sum_{i=1}^{m}\|\boldsymbol{a}_i\|_2 = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$$

Matrix Norms

- $l_{2,1}$ norm can be generalized to the norm. $p, q \geq 1$, defined by

$$\|\boldsymbol{A}\|_{p,q} = \left(\sum_{i=1}^{m}\left(\sum_{j=1}^{n}|a_{ij}|^{p}\right)^{q/p}\right)^{1/q}$$

- $p = 2, q = 1$

$$\|\boldsymbol{A}\|_{2,1} = \sum_{i=1}^{m}\|\boldsymbol{a}_i\|_2 = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$$

Matrices

- **Determinant:** $|\boldsymbol{A}|$, $\boldsymbol{A}_{n\times n}$ is a square matrix

$$|\boldsymbol{A}| = \sum_{k=1}^{d} a_{ik}|\boldsymbol{A}_{ik}|(-1)^{k+i}$$

- **Minor matrix $\boldsymbol{A}_{ik}$** is formed by removing the i^{th}row and the k^{th} column of $\boldsymbol{A}$.

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 3 & 0 & 5 \\ -1 & 9 & 11 \end{pmatrix} \qquad \boldsymbol{A}_{23} = ?$$

Matrices

- **Determinant:** $|\boldsymbol{A}|$, $\boldsymbol{A}_{n\times n}$ is a square matrix

$$|\boldsymbol{A}| = \sum_{k=1}^{d} a_{ik}|\boldsymbol{A}_{ik}|(-1)^{k+i}$$

- **Minor matrix $\boldsymbol{A}_{ik}$** is formed by removing the i^{th} row and the k^{th} column of $\boldsymbol{A}$.

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 3 & 0 & 5 \\ -1 & 9 & 11 \end{pmatrix} \quad \boldsymbol{A}_{23} = \begin{pmatrix} 1 & 4 \\ -1 & 9 \end{pmatrix}$$

Matrices

- **Determinant: $|\boldsymbol{A}|$, $\boldsymbol{A}_{n\times n}$** is a square matrix

$$|\boldsymbol{A}| = \sum_{k=1}^{d} a_{ik}|\boldsymbol{A}_{ik}|(-1)^{k+i}$$

- **Minor matrix $\boldsymbol{A}_{ik}$** is formed by removing the i^{th} row and the k^{th} column of $\boldsymbol{A}$.

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 3 & 0 & 5 \\ -1 & 9 & 11 \end{pmatrix} \quad \boldsymbol{A}_{23} = \begin{pmatrix} 1 & 4 \\ -1 & 9 \end{pmatrix}$$

- **Properties**
  - $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$, where $\boldsymbol{A}_{n\times n}$ and $\boldsymbol{B}_{n\times n}$
  - $|\boldsymbol{A}^{-1}| = 1/|\boldsymbol{A}|$, $|\boldsymbol{A}^{T}| = |\boldsymbol{A}|$
  - Singular or non-singular
    - A singular matrix has a zero determinant
    - A non-singular matrix has a non-zero determinant

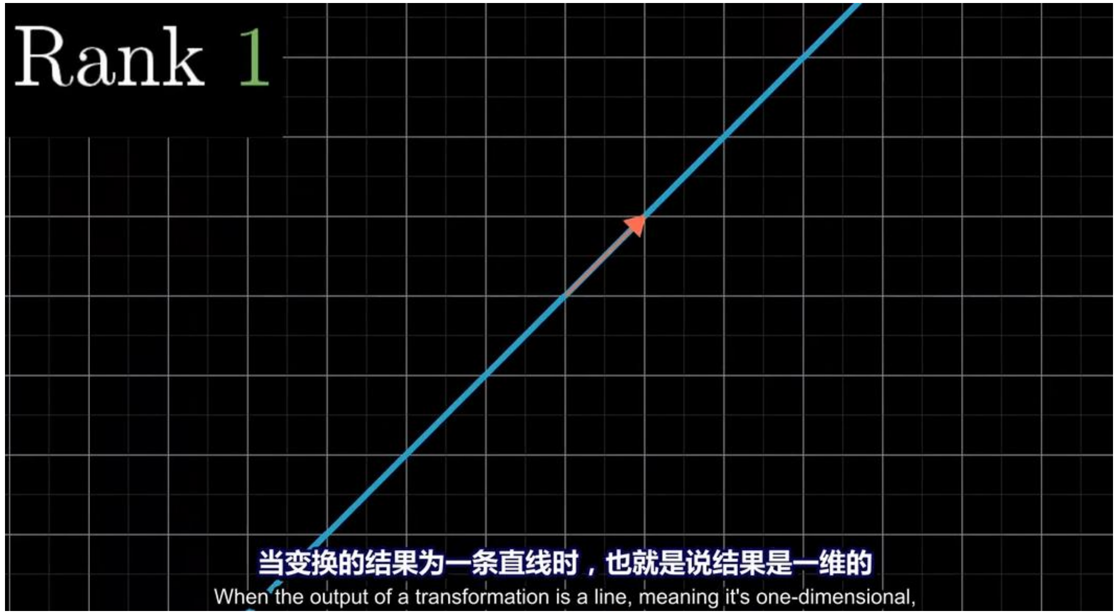
Matrices

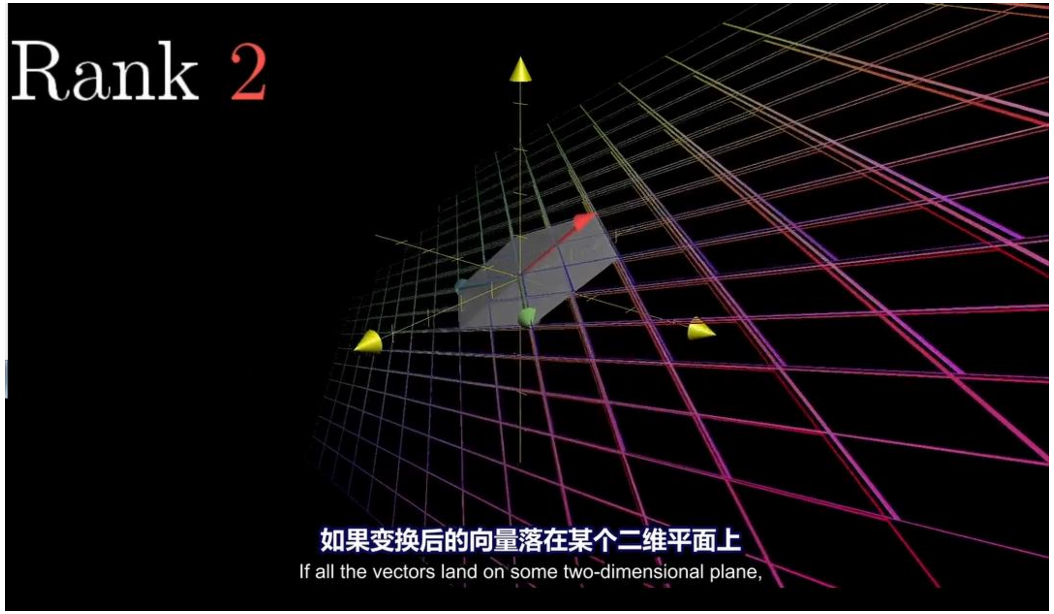
- **rank:** the number of linearly independent rows (or columns).

Matrices

- **rank:** the number of linearly independent rows (or columns).
- More intuitive explanation:

rank ⟷ **Number of dimensions in the output**





Matrices

- **rank:** the number of linearly independent rows (or columns)
- **Note:**
- Given a matrix $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{B}_{n\times k}$, then we have

$$0 \le rank(\boldsymbol{A}) \le min(m, n)$$

- Full rank: $rank(\boldsymbol{A}) = min(m, n)$

Matrices

- **rank:** the number of linearly independent rows (or columns)
- **Note:**
- Given a matrix $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{B}_{n\times k}$, then we have

$$0 \leq rank(\boldsymbol{A}) \leq min(m, n)$$

- Full rank: $rank(\boldsymbol{A}) = min(m, n)$
- $rank(\boldsymbol{AB}) = min(rank(\boldsymbol{A}), rank(\boldsymbol{B}))$
- $rank(\boldsymbol{A} + \boldsymbol{B}) \leq rank(\boldsymbol{A}) + rank(\boldsymbol{B})$
- $rank(\boldsymbol{A}^T\boldsymbol{A}) = rank(\boldsymbol{A}\boldsymbol{A}^T) = rank(\boldsymbol{A}) = rank(\boldsymbol{A}^T)$

Matrices

- **rank:** the number of linearly independent rows (or columns)
- Note:
- Given a matrix $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{B}_{n\times k}$, then we have

$$0 \leq rank(\boldsymbol{A}) \leq min(m,n)$$

- Full rank: $rank(\boldsymbol{A}) = min(m,n)$
- $rank(\boldsymbol{AB}) = min(rank(\boldsymbol{A}), rank(\boldsymbol{B}))$
- $rank(\boldsymbol{A}+\boldsymbol{B}) \leq rank(\boldsymbol{A}) + rank(\boldsymbol{B})$
- $rank(\boldsymbol{A}^T\boldsymbol{A}) = rank(\boldsymbol{A}\boldsymbol{A}^T) = rank(\boldsymbol{A}) = rank(\boldsymbol{A}^T)$
- Given a square matrix $\boldsymbol{A}_{n\times n}$,
  - Singular : $rank(\boldsymbol{A}) < n$
  - Non-singular: $rank(\boldsymbol{A}) = n, rank(\boldsymbol{A}^{-1}) = rank(\boldsymbol{A})$

Linear Transformations

- Mapping from vector space $\boldsymbol{X} \in \mathbb{R}^N$ to vector space $\boldsymbol{Y} \in \mathbb{R}^M$, represented by a matrix:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1N} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \ddots & \\ a_{M1} & a_{M2} & a_{M3} & & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

$\boldsymbol{V}, \boldsymbol{W}$ are vector spaces. $f: \boldsymbol{V} \rightarrow \boldsymbol{W}$ is said to be *a linear map* if for any vectors $\boldsymbol{u}$ and $\boldsymbol{v} \in \boldsymbol{V}$ and any scalar c the following is satisfied:

- $f(\boldsymbol{u}+\boldsymbol{v}) = f(\boldsymbol{u}) + f(\boldsymbol{v})$
- $f(c\boldsymbol{u}) = cf(\boldsymbol{u})$

- **Note**
  - The dimensionality of the two spaces does not need to be the same.
  - For machine learning, typically $M < N$, i.e., project onto a lower-dimensionality space.

Eigenvectors and Eigenvalues

- **Definition:** $\boldsymbol{v}$ is an eigenvector of matrix $\boldsymbol{A} \in \mathbb{R}^{m*m}$ if there exists a scalar λ, such that:

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \qquad \begin{cases} \boldsymbol{v}\text{: an eigenvector} \\ \lambda\text{: the corresponding eigenvalue} \end{cases}$$

- **Computation**

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \qquad (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$$

Trivial solution: $\boldsymbol{v} = \boldsymbol{0}$

Non-trivial solution: $\boldsymbol{v} \neq \boldsymbol{0} \quad \overset{|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|}{\Rightarrow} \quad |\boldsymbol{A} - \lambda\boldsymbol{I}| = \boldsymbol{0}$

Eigenvectors and Eigenvalues

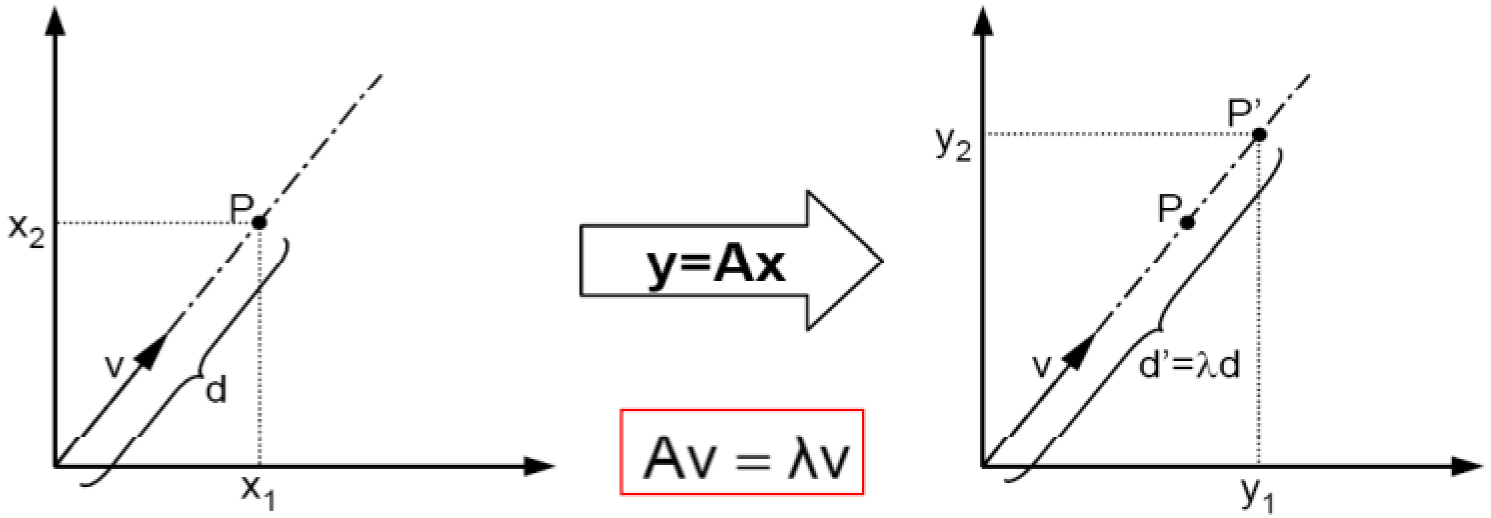
- **Definition:** $\boldsymbol{v}$ is an eigenvector of matrix $\boldsymbol{A} \in \mathbb{R}^{m*m}$ if there exists a scalar λ, such that:

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v} \quad \begin{cases} \boldsymbol{v}\text{: an eigenvector} \\ \lambda\text{: the corresponding eigenvalue} \end{cases}$$

- **Note**
  - $tr(\boldsymbol{A}) = \sum_i \lambda_i$
  - $|\boldsymbol{A}| = \prod_i \lambda_i$
  - If λ is an eigenvalue of the matrix $\boldsymbol{A}$, then λ^2 is an eigenvalue of $\boldsymbol{A}^2$. $(\boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{A})$
  - If λ is an eigenvalue of the matrix $\boldsymbol{A}$, prove then λ is an eigenvalue of $\boldsymbol{A}^T$.

Eigenvectors and Eigenvalues

- Intepretation: an eigenvector represents an invariant direction in the vector space.
  - Any point lying on the direction defined by $\boldsymbol{v}$ remains on that direction.
  - Its magnitude is multiplies by the corresponding eigenvalue λ



Similar Matrix

- **Similar Matrix**
  - In linear algebra, two n-by-n matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are called similar if $\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, for some invertible n-by-n matrix $\boldsymbol{P}$.

- **Properties**
  - $rank(\boldsymbol{A}) = rank(\boldsymbol{B})$
  - $tr(\boldsymbol{A}) = tr(\boldsymbol{B})$
  - $|\boldsymbol{A}| = |\boldsymbol{B}|$
  - $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same eigenvalues!

Similar Matrix

- **Similar Matrix**
  - In linear algebra, two n-by-n matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are called similar if $\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, for some invertible n-by-n matrix $\boldsymbol{P}$.

- **Properties**
  - $rank(\boldsymbol{A}) = rank(\boldsymbol{B})$
  - $tr(\boldsymbol{A}) = tr(\boldsymbol{B})$
  - $|\boldsymbol{A}| = |\boldsymbol{B}|$
  - $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same eigenvalues! **WHY?**

Assume that $\boldsymbol{Bv} = \lambda\boldsymbol{v}$, then $\boldsymbol{Bv} = \boldsymbol{P}^{-1}\boldsymbol{APv}$.
Multiply $\boldsymbol{P}$ on both sides, $\boldsymbol{PBv} = \boldsymbol{APv} \Rightarrow \boldsymbol{P}\lambda\boldsymbol{v} = \boldsymbol{APv} \Rightarrow \lambda(\boldsymbol{Pv}) = \boldsymbol{A}(\boldsymbol{Pv})$.
This means that λ is also the eigenvalue of $\boldsymbol{A}$ with the eigenvector $\boldsymbol{Pv}$.

Eigenvalue Decomposition

- Given a **square** matrix with m linearly **independent** eigenvectors $\boldsymbol{A} \in \mathbb{R}^{m*m}$, we have an eigenvalue decomposition

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

- <u>Note</u>
  - Columns of $\boldsymbol{V}$ are eigenvectors of $\boldsymbol{A}$
  - Diagonal elements of $\boldsymbol{\Lambda}$ are eigenvalues of A

$$\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \dots, \lambda_m), \lambda_i \geq \lambda_{i+1}$$

Eigenvalue Decomposition

- **Note**
- If $\boldsymbol{A}$ is non-singular
  - All eigenvalues are non-zero (WHY?)

Eigenvalue Decomposition

- **Note**
- If $\boldsymbol{A}$ is non-singular
  - All eigenvalues are non-zero (WHY?) $|\boldsymbol{A}| = \prod_i \lambda_i$

Eigenvalue Decomposition

- **Note**
- If $\boldsymbol{A}$ is non-singular
  - All eigenvalues are non-zero (WHY?) $|\boldsymbol{A}| = \prod_i \lambda_i$
- If A is real and symmetric
  - All eigenvalues are *real*.

$$\text{If } |\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \text{ and } \boldsymbol{A} = \boldsymbol{A}^T \Rightarrow \lambda \in \mathbb{R}$$

  - The eigenvectors for distinct eigenvalues are *orthogonal*.

$$\boldsymbol{A}\boldsymbol{v}_{\{1,2\}} = \lambda_{\{1,2\}}\boldsymbol{v}_{\{1,2\}} \text{ and } \lambda_1 \neq \lambda_2 \Rightarrow \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = 0$$

Eigenvalue Decomposition

- **Note**
- If $\boldsymbol{A}$ is non-singular
  - All eigenvalues are non-zero (WHY?) $|\boldsymbol{A}| = \prod_i \lambda_i$
- If A is real and symmetric
  - All eigenvalues are *real*.

$$\text{If } |\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \text{ and } \boldsymbol{A} = \boldsymbol{A}^{\boldsymbol{T}} \Rightarrow \lambda \in \mathbb{R}$$

  - The eigenvectors for distinct eigenvalues are *orthogonal*.

$$\boldsymbol{A}\boldsymbol{v}_{\{1,2\}} = \lambda_{\{1,2\}}\boldsymbol{v}_{\{1,2\}} \text{ and } \lambda_1 \neq \lambda_2 \Rightarrow \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = 0$$

- If A is positive definite, then all eigenvalues are *positive*.

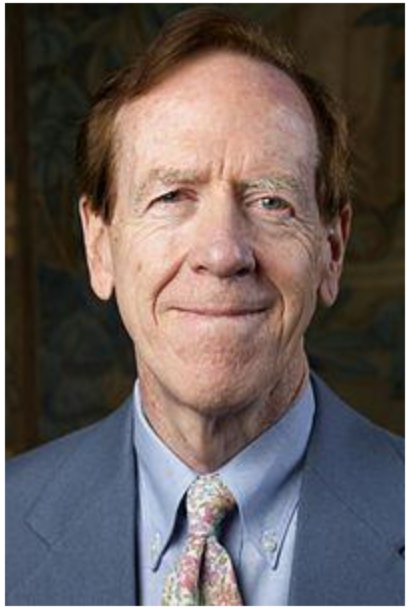
$$\forall \boldsymbol{w} \in \mathbb{R}^n, \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{A}\boldsymbol{w} > 0, \text{ then if } \boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v} \Rightarrow \lambda > 0$$

Singular Value Decomposition (SVD)

- "The highpoint of linear algebra"--- *Gillert Strang*

Awards and honors [edit]

- Rhodes Scholar (1955)
- Alfred P. Sloan Fellow (1966–1967)
- Chauvenet Prize, Mathematical Association of America (1976)
- Honorary Professor, Xian Jiaotong University, China (1980)
- American Academy of Arts and Sciences (1985)
- Honorary Fellow, Balliol College, Oxford University (1999)
- Honorary Member, Irish Mathematical Society (2002)
- Award for Distinguished Service to the Profession, Society for Industrial and Applied Mathematics (2003)
- Lester R. Ford Award (2005)[3]
- Von Neumann Medal, US Association for Computational Mechanics (2005)
- Haimo Prize, Mathematical Association of America (2007)[4]
- Su Buchin Prize, International Congress (ICIAM, 2007)
- Henrici Prize (2007)
- National Academy of Sciences (2009)
- Doctor Honoris Causa, University of Toulouse (2010)
- Fellow of the American Mathematical Society (2012)[5]
- Doctor Honoris Causa, Aalborg University (2013)
- Fellow of the Society for Industrial and Applied Mathematics (2009) [6]



1934~

Singular Value Decomposition (SVD)

- **“The highpoint of linear algebra”---** *Gillert Strang*
- Any $m \times n$ matrix $\boldsymbol{A}$ of rank r can be decomposed into: $\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$
  - For $m > n$

$$
\begin{array}{cccc}
 & \boldsymbol{U}_{m\times m} & \boldsymbol{\Sigma}_{m\times n} & \boldsymbol{V}_{n\times n} \\
\mathbf{A} = &
\begin{bmatrix}
\vdots & & \vdots & \vdots & & \vdots \\
\mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m \\
\vdots & & \vdots & \vdots & & \vdots
\end{bmatrix}
&
\begin{bmatrix}
\sigma_1 & & & & & \\
 & \ddots & & & & \\
 & & \sigma_r & & & \\
 & & & 0 & & \\
 & & & & \ddots & \\
 & & & & & 0 \\
0 & \dots & \dots & \dots & \dots & 0 \\
\vdots & & & & & \vdots \\
0 & \dots & \dots & \dots & \dots & 0
\end{bmatrix}
&
\begin{bmatrix}
\dots & \mathbf{v}_1^\top & \dots \\
 & \vdots & \\
\dots & \mathbf{v}_r^\top & \dots \\
\dots & \mathbf{v}_{r+1}^\top & \dots \\
 & \vdots & \\
\dots & \mathbf{v}_n^\top & \dots
\end{bmatrix}
\end{array}
$$

Singular Value Decomposition (SVD)

- **"The highpoint of linear algebra"---** *Gillert Strang*
- Any $m \times n$ matrix $\boldsymbol{A}$ of rank r can be decomposed into: $\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$
  - For $m > n$

$$\mathbf{A} = \overset{\boldsymbol{U}_{m\times m}}{\begin{bmatrix} \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \\ \vdots & & \vdots & \vdots & & \vdots \end{bmatrix}} \overset{\boldsymbol{\Sigma}_{m\times n}}{\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}} \overset{\boldsymbol{V}_{n\times n}}{\begin{bmatrix} \cdots & \mathbf{v}_1^\top & \cdots \\ & \vdots & \\ \cdots & \mathbf{v}_r^\top & \cdots \\ \cdots & \mathbf{v}_{r+1}^\top & \cdots \\ & \vdots & \\ \cdots & \mathbf{v}_n^\top & \cdots \end{bmatrix}}$$

- Each of these matrices is defined to have a special structure：
  - The columns of $\boldsymbol{U}$ (i.e., left singular vectors) are orthogonal, i.e., $\boldsymbol{UU^T} = \boldsymbol{I}_{m\times m}$.
  - The columns of $\boldsymbol{V}$ (i.e., right singular vectors) are orthogonal, i.e., $\boldsymbol{VV^T} = \boldsymbol{I}_{n\times n}$.
  - We arrange the singular values as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Singular Value Decomposition (SVD)

- **"The highpoint of linear algebra"---** *Gillert Strang*
- Any $m \times n$ matrix $\boldsymbol{A}$ of rank r can be decomposed into: $\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$
  - For $m > n$ $\quad \boldsymbol{U}_{m\times m} \quad \boldsymbol{\Sigma}_{m\times n} \quad \boldsymbol{V}_{n\times n}$

$$\mathbf{A} = \begin{bmatrix} \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \\ \vdots & & \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} \cdots & \mathbf{v}_1^\top & \cdots \\ & \vdots & \\ \cdots & \mathbf{v}_r^\top & \cdots \\ \cdots & \mathbf{v}_{r+1}^\top & \cdots \\ & \vdots & \\ \cdots & \mathbf{v}_n^\top & \cdots \end{bmatrix}$$

- Each of these matrices is defined to have a special structure：
  - The columns of $\boldsymbol{U}$ (i.e., left singular vectors) are eigenvectors of $\boldsymbol{AA^T}$.
  - The columns of $\boldsymbol{V}$ (i.e., right singular vectors) are eigenvectors of $\boldsymbol{A^T A}$.
  - Eigenvalues $\lambda_1, \dots, \lambda_r$ of $\boldsymbol{AA^T}$ are the eigenvalues of $\boldsymbol{A^T A}$.
  - Singular value $\sigma_i = \sqrt{\lambda_i}$.

Singular Value Decomposition (SVD)

- **"The highpoint of linear algebra"---** *Gillert Strang*
- Any $m \times n$ matrix $\boldsymbol{A}$ of rank r can be decomposed into: $\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$
  - For $m > n$ $\quad \boldsymbol{U}_{m\times m} \quad \boldsymbol{\Sigma}_{m\times n} \quad \boldsymbol{V}_{n\times n}$

$$
\mathbf{A} = \begin{bmatrix} \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \\ \vdots & & \vdots & \vdots & & \vdots \end{bmatrix}
\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix}
\begin{bmatrix} \cdots & \mathbf{v}_1^\top & \cdots \\ & \vdots & \\ \cdots & \mathbf{v}_r^\top & \cdots \\ \cdots & \mathbf{v}_{r+1}^\top & \cdots \\ & \vdots & \\ \cdots & \mathbf{v}_n^\top & \cdots \end{bmatrix}
$$

- Each of these matrices is defined to have a special struc
  - The columns of $\boldsymbol{U}$ (i.e., left singular vectors) are eige
  - The columns of $\boldsymbol{V}$ (i.e., right singular vectors) are eig
  - Eigenvalues $\lambda_1, \ldots, \lambda_r$ of $\boldsymbol{AA^T}$ are the eigenvalues of
  - Singular value $\sigma_i = \sqrt{\lambda_i}$.

Singular Value Decomposition (SVD)

- **"The highpoint of linear algebra"---** *Gillert Strang*
- Any $m \times n$ matrix $\boldsymbol{A}$ of rank r can be decomposed into: $\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$
  - For $m > n$ — $\boldsymbol{U}_{m\times m}$ — $\boldsymbol{\Sigma}_{m\times n}$ — $\boldsymbol{V}_{n\times n}$

σ_1 $\ddots$ $[\cdots \; \mathbf{v}_1^{\mathrm{T}} \; \cdots]$

$$\boldsymbol{A^T A} = (\boldsymbol{U\Sigma V^T})^{\boldsymbol{T}}\boldsymbol{U\Sigma V^T} = \boldsymbol{V\Sigma^2 V^T}$$

- Eigenvalues $\lambda_1, \dots, \lambda_r$ of $\boldsymbol{AA^T}$ are the eigenvalues of
- Singular value $\sigma_i = \sqrt{\lambda_i}$.

Singular Value Decomposition (SVD)

- **"The highpoint of linear algebra"---** *Gillert Strang*
- Any $m \times n$ matrix $\boldsymbol{A}$ of rank r can be decomposed into: $\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$
  - ➢ For $m > n$ $\quad \boldsymbol{U}_{m\times m} \quad \boldsymbol{\Sigma}_{m\times n} \quad \boldsymbol{V}_{n\times n}$

σ_1 $\ddots$ $[\cdots \; \mathbf{v}_1^T \; \cdots]$

$$\boldsymbol{A^T A} = (\boldsymbol{U\Sigma V^T})^{\boldsymbol{T}} \boldsymbol{U\Sigma V^T} = \boldsymbol{V\Sigma^2 V^T}$$

This is the eigen-decomposition of $\boldsymbol{A^T A}$.

$\boldsymbol{V}$ is the eigenvector matrix of $\boldsymbol{A^T A}$, and $\boldsymbol{\Sigma^2}$ is the eigenvalue matrix of $\boldsymbol{A^T A}$, i.e., singular values are positive square roots of eigenvalues.

- Eigenvalues $\lambda_1, \dots, \lambda_r$ of $\boldsymbol{AA^T}$ are the eigenvalues of
- Singular value $\sigma_i = \sqrt{\lambda_i}$.

Singular Value Decomposition (SVD)

- "The highpoint of linear algebra"--- *Gillert Strang*
- Any $m \times n$ matrix $\boldsymbol{A}$ of rank r can be decomposed into: $\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$
  - For $m > n$ $\qquad \boldsymbol{U}_{m\times m} \qquad \boldsymbol{\Sigma}_{m\times n} \qquad \boldsymbol{V}_{n\times n}$

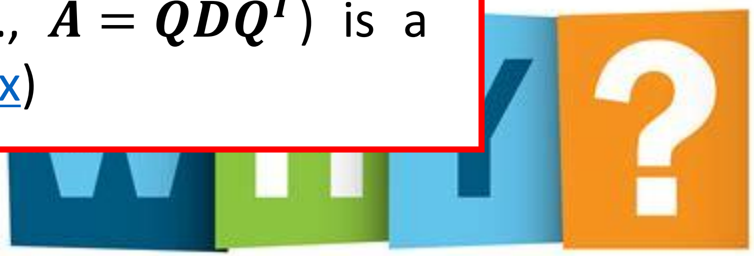
σ_1 $\ddots$ $[\cdots \ \mathbf{v}_1^T \ \cdots]$

$$\boldsymbol{A^T A} = (\boldsymbol{U\Sigma V^T})^T \boldsymbol{U\Sigma V^T} = \boldsymbol{V\Sigma^2 V^T}$$

This is the eigen-decomposition of $\boldsymbol{A^T A}$.

$\boldsymbol{V}$ is the eigenvector matrix of $\boldsymbol{A^T A}$, and $\boldsymbol{\Sigma^2}$ is the eigenvalue matrix of $\boldsymbol{A^T A}$, i.e., singular values are positive square roots of eigenvalues.

Theorem1: Every symmetric matrix $\boldsymbol{M}$ is orthogonally diagonalizable, i.e., there exists an orthogonal matrix $\boldsymbol{Q}$ (i.e., $\boldsymbol{Q^T} = \boldsymbol{Q^{-1}}$) such that $\boldsymbol{Q^T A Q} = \boldsymbol{D}$ (i.e., $\boldsymbol{A} = \boldsymbol{QDQ^T}$) is a diagonal matrix. (https://en.wikipedia.org/wiki/Diagonalizable_matrix)



- Eigenvalues $\lambda_1, \dots, \lambda_r$ of $\boldsymbol{AA^T}$ are the eigenvalues of
- Singular value $\sigma_i = \sqrt{\lambda_i}$.

Singular Value Decomposition (SVD)

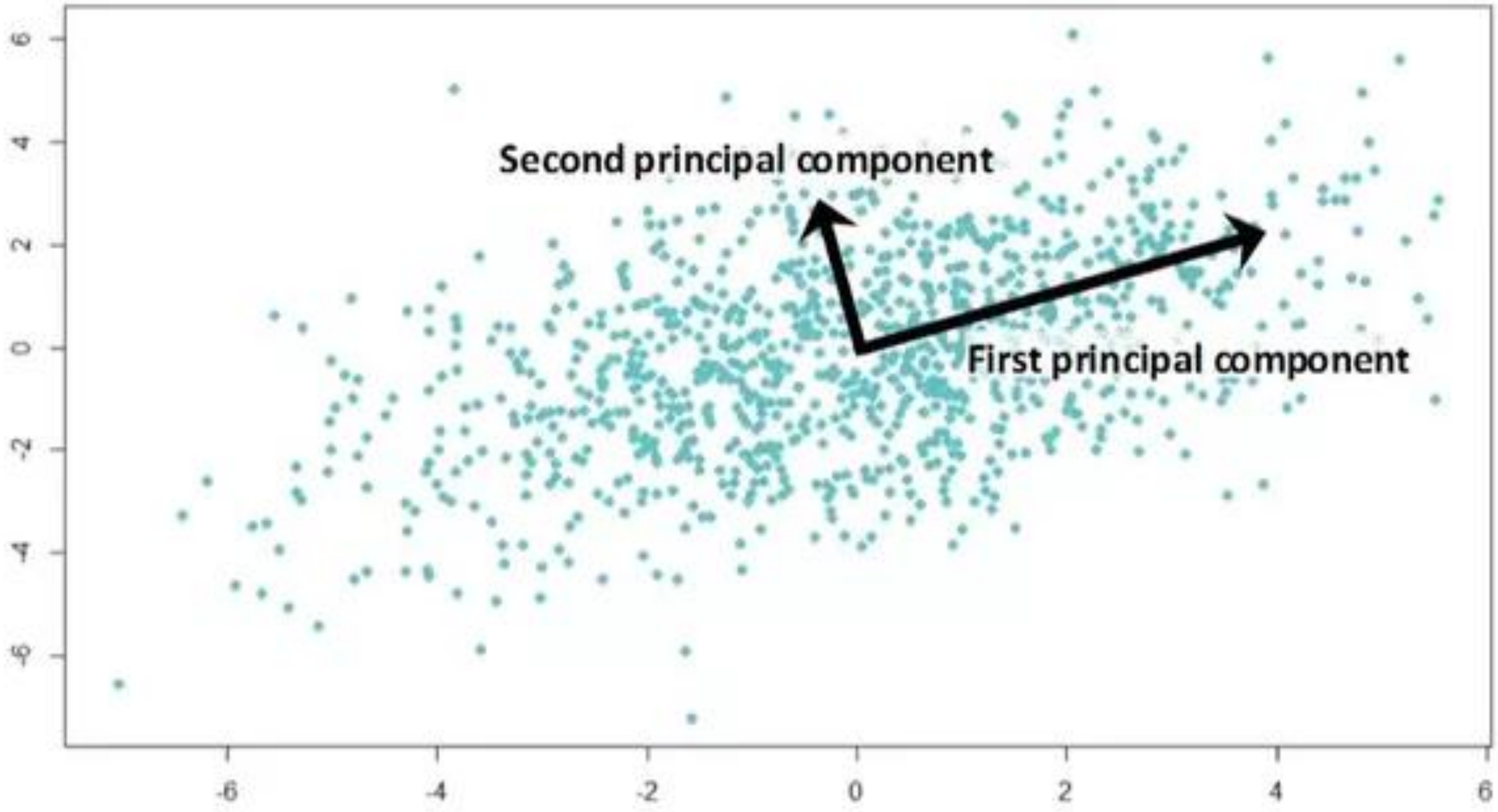
- For $m > n$

$$
\mathbf{A} = \overset{\boldsymbol{U}_{m\times m}}{\begin{bmatrix} \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m \\ \vdots & & \vdots & \vdots & & \vdots \end{bmatrix}}
\overset{\boldsymbol{\Sigma}_{m\times n}}{\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{bmatrix}}
\overset{\boldsymbol{V}_{n\times n}}{\begin{bmatrix} \dots & \mathbf{v}_1^\top & \dots \\ & \vdots & \\ \dots & \mathbf{v}_r^\top & \dots \\ \dots & \mathbf{v}_{r+1}^\top & \dots \\ & \vdots & \\ \dots & \mathbf{v}_n^\top & \dots \end{bmatrix}}
$$

- Arrange $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, $r =$rank($\boldsymbol{A}$)
- Economy version $\boldsymbol{A} = \underset{m\times r}{\underline{\boldsymbol{U}_r}}\,\underset{r\times r}{\underline{\boldsymbol{\Sigma}_r}}\,\underset{r\times n}{\underline{\boldsymbol{V}_r}}{}^{\boldsymbol{T}}$

$$\boldsymbol{\Sigma}_r = \mathrm{diag}(\sigma_1, \dots, \sigma_r)$$

SVD--Principal Component Analysis (PCA)



SVD--Latent Semantic Indexing (LSI)

Index Words	Titles								
	T1	T2	T3	T4	T5	T6	T7	T8	T9
book			1	1					
dads						1			1
dummies		1						1	
estate							1		1
guide	1					1			
investing	1	1	1	1	1	1	1	1	1
market	1		1						
real							1		1
rich						2			1
stock	1		1					1	
value				1	1				

A

book	0.15	-0.27	0.04
dads	0.24	0.38	-0.09
dummies	0.13	-0.17	0.07
estate	0.18	0.19	0.45
guide	0.22	0.09	-0.46
investing	0.74	-0.21	0.21
market	0.18	-0.30	-0.28
real	0.18	0.19	0.45
rich	0.36	0.59	-0.34
stock	0.25	-0.42	-0.28
value	0.12	-0.14	0.23

*

3.91	0	0
0	2.61	0
0	0	2.00

*

T1	T2	T3	T4	T5	T6	T7	T8	T9
0.35	0.22	0.34	0.26	0.22	0.49	0.28	0.29	0.44
-0.32	-0.15	-0.46	-0.24	-0.14	0.55	0.07	-0.31	0.44
-0.41	0.14	-0.16	0.25	0.22	-0.51	0.55	0.00	0.34

U_k Σ_k ${V_k}^T$

$k = 3$

Derivatives

Differentiate w.r.t	scalar	vector	matrix
scalar	scalar	vector	matrix
vector	vector	matrix	
matrix	matrix		

- **scalar –scalar:** e.g., $\frac{d}{dx}x^2 = 2$
- **scalar-vector:** e.g., $f(\boldsymbol{x})$ is a scalar function of vector $\boldsymbol{x}$

$$\mathrm{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \qquad \frac{df}{d\mathbf{x}} = \begin{bmatrix} \frac{\sigma f}{\sigma x_1} \\ \vdots \\ \frac{\sigma f}{\sigma x_d} \end{bmatrix}$$

- **scalar-matrix:** $f(\boldsymbol{A})$ is a scalar function and $m \times n$ matrix $\boldsymbol{A}$

$$\frac{df}{d\mathbf{A}} = \begin{bmatrix} \frac{\sigma f}{\sigma a_{11}} & \cdots & \frac{\sigma f}{\sigma a_{1d}} \\ \vdots & \ddots & \vdots \\ \frac{\sigma f}{\sigma a_{m1}} & \cdots & \frac{\sigma f}{\sigma a_{mn}} \end{bmatrix}$$

Commonly Used Derivatives

- $\frac{d}{dx}(\boldsymbol{Ax}) = \boldsymbol{A^T}$

- $\frac{d\boldsymbol{x}}{d\boldsymbol{x}} = \boldsymbol{I}$

- $\frac{d\boldsymbol{y^T x}}{d\boldsymbol{x}} = \frac{d\boldsymbol{x^T y}}{d\boldsymbol{x}} = \boldsymbol{y}$

- $\frac{d}{dx}(\boldsymbol{x^T A x}) = \begin{cases} (\boldsymbol{A} + \boldsymbol{A^T})\boldsymbol{x} & \text{If } \mathbf{A} \text{ square} \\ \boldsymbol{2Ax} & \text{If } \mathbf{A} \text{ symmetric} \end{cases}$

https://en.wikipedia.org/wiki/Matrix_calculus

Summary

- **Vectors**
  - **Products and norms**
    - Vector norms l_1-$norm$, l_2-$norm$, l_∞-$norm$
  - **Linear dependence and independence**
    - $a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n = 0 \Rightarrow a_k = 0 \quad \forall k$
  - **Vector spaces and basis**
    - $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ are **independent** $\Leftrightarrow$ they form a basis
    - Orthonormal and Orthogonal

Summary

- **Matrices**
  - **Trace**
    - $tr(\boldsymbol{ABC}) = tr(\boldsymbol{BCA}) = tr(\boldsymbol{CAB})$ (Cyclic property)
  - **Determinant**
    - A singular matrix has a zero determinant
    - $|\boldsymbol{A}^{-1}| = 1/|\boldsymbol{A}|$
    - $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$
  - **rank:** the number of linearly independent rows (or columns);
    the number of dimensions in the output
  - **Positive definite and semi-positive definite matrix**
    - $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$ (★ all eigenvalues are *positive*)
    - $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$

Summary

- **Matrices**
  - **(Real) Symmetric matrix**
    - All eigenvalues are *real*.
    - The eigenvectors for distinct eigenvalues are *orthogonal*.
  - **Norms**
    - $l_{2,1}$-$norm$, Frobenius norm ★ $\|\boldsymbol{A}\|_F = (tr(\boldsymbol{A}^T\boldsymbol{A}))^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$
  - **Eigenvalues and eigenvectors**
    - Eigen-decomposition ($\boldsymbol{AV} = \boldsymbol{V\Lambda}$)
      - ➢ $tr(\boldsymbol{A}) = \sum_i \lambda_i$
      - ➢ $|\boldsymbol{A}| = \prod_i \lambda_i$
    - Singular Value Decomposition ($\boldsymbol{A} = \boldsymbol{U\Sigma V^T}$)