Linear Transformations, Matrices & Determinant

Thomas Schnake, Naima Elosegui Borras, Tom Kaufmann

Machine Learning Group

May 28, 2024

Table of Contents

Linear Transformation

Definition. Let V and W be F-vector spaces. A transformation $f\colon V \to W$ is called **linear** if it satisfies

1. $f(v+\omega) = f(v) + f(\omega)$ and
2. $f(\lambda v) = \lambda f(v)$

$\forall v, \omega \in V$ and $\lambda \in F$

Existence & Uniqueness

Theorem. Let V and W be real vector spaces and let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a basis of V. Then for *any* n-tuple $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ of vectors in W there exists exactly one linear transformation

$$f\colon V \to W \qquad \text{s.t.} \quad f(\boldsymbol{v}_i) = \boldsymbol{w}_i \quad i = 1, \ldots, n.$$

Table of Contents

Matrix

Definition. A real $m \times n$ **matrix** is a rectangular array of $m \cdot n$ real numbers $a_{11}, \ldots, a_{mn}$ arranged as

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

The a_{ij} are called the **coefficients** of the matrix.

Examples

- identity matrix I
- $g : \mathbb{R}^2 \to \mathbb{R}^2$ rotation by 90°
- $g : \mathbb{R}^2 \to \mathbb{R}^2$ reflection along the x-axis
- $g : \mathbb{R}^2 \to \mathbb{R}^2$ projection on the line through $(0, 0)$ and $(1, 1)$

Matrix arithmetic

- addition: $C = A + B$ with $c_{ij} = a_{ij} + b_{ij}$
- scalar multiplication: $C = \lambda A$ with $c_{ij} = \lambda a_{ij}$
- multiplication: $C = AB$ with

$$c_{ik} = \sum_j a_{ij} b_{jk} \qquad \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

where $A \in \mathbb{R}^{l \times m}$, $B \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{R}^{l \times n}$

Matrix transpose

Definition. Let $A \in \mathbb{R}^{m \times n}$ be a real matrix. The **transpose** of A is defined as

$$A^\top_{ji} = A_{ij} \quad \text{for} \begin{cases} i = 1, \ldots, m \\ j = 1, \ldots, n. \end{cases}$$

The following rules hold:

1. $(A + B)^\top = A^\top + B^\top$
2. $(\lambda \cdot A)^\top = \lambda \cdot A^\top$
3. $(A \cdot C)^\top = C^\top \cdot A^\top$

Rank

Definition: The **rank** $\mathrm{rank}(A)$ of a matrix A is the maximum number of linearly independent rows of a matrix.

Theorem. The row rank is equal to the column rank:

$$\forall A \in \mathbb{R}^{m\times n}:\ \mathrm{rank}(A) = \mathrm{rank}(A^\top)$$

The rows of $A \in \mathbb{R}^{m\times n}$ span a subspace of $\mathcal{V} \subseteq \mathbb{R}^n$ with dimensionality $\mathrm{rank}(A)$.

Matrices as Linear Transformations

Any linear map $f \colon \mathbb{R}^n \to \mathbb{R}^m$ can be defined by a unique matrix $A \in \mathbb{R}^{m \times n}$ via

$$f(\boldsymbol{x}) = A\boldsymbol{x}$$

where the dimension of the image of f is $\mathrm{rank}(A)$.

Note: Let $\boldsymbol{z} = A\boldsymbol{x}$. In machine learning we usually say "the dimension of $\boldsymbol{z}$ is m" ignoring the fact that the rows of A could be linearly dependent ($\mathrm{rank}(A) < m$).

Matrix inverse

Definition: A square matrix $A \in \mathbb{R}^{n \times n}$ is **invertible** (**regular**, **non-singular**) if there exists a matrix A^{-1} such that

$$AA^{-1} = A^{-1}A = I_n.$$

The following statements hold:

- A is invertible if and only if $\operatorname{rank}(A) = n$
- $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$
- $(A^\top)^{-1} = (A^{-1})^\top$
- $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Trace

Definition. The **trace** $\mathrm{tr}(A)$ of a square matrix A is the sum of the elements on its main diagonal.

Theorem. For $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times n}$ it holds that

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

Special types of matrices

Let $A \in \mathbb{R}^{n \times n}$ be a real square matrix.

- A is **orthogonal** if $AA^\top = A^\top A = I_n$. Orthogonal matrices represent reflections and rotations of a space.
- A is **symmetric** if $A = A^\top$
- A is **anti-symmetric** if $A = -A^\top$
- A is **diagonal** if all of its coefficients apart from the main diagonal are 0:
  $\forall i, j: \; i \neq j \;\Rightarrow\; a_{ij} = 0$

Table of Contents

Determinant

The determinant of a square matrix is a single number. What does it represent?

- absolute value: volume of the parallelotope spanned by the column or row vectors of the matrix
- sign: orientation of the parallelotope

Determinant

Theorem and Definition: There exists exactly one mapping $\det : \mathbb{R}^{n \times n} \to \mathbb{R},\ A \mapsto \det(A)$ that satisfies

1. $\det$ is linear in every column
2. If $\operatorname{rank}(A) < n$ then $\det(A) = 0$
3. $\det(I_n) = 1$

This mapping is called the **determinant**.

Determinant - useful facts

Let $A, B \in \mathbb{R}^{n \times n}$ be two square matrices and $\lambda \in \mathbb{R}$:

- If A is triangular, its determinant is the product of the coefficients on its main diagonal
- A is invertible if and only if $\det A \neq 0$
- $\det(AB) = \det(A)\det(B)$
- $\det(A^{-1}) = (\det(A))^{-1}$
- $\det(A) = \det(A^\top)$
- $\det(\lambda A) = \lambda^n \det(A)$
- For 2×2 matrices: $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$

Table of Contents

Projections

Definition. Let V be a real vector space. A linear transformation $P\colon V \to V$ is a **projection** if it holds that $P \circ P = P$.

Example: projection of a 3-dimensional object onto a two-dimensional subspace.

Definition. Let U be a vector subspace of V and let $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ be an orthonormal basis of U. Then

$$P_U(\boldsymbol{x}) := \sum_{i=1}^{n} \langle \boldsymbol{x}, \mathbf{u}_i \rangle \mathbf{u}_i$$

defines an **orthogonal projection** onto U.

Table of Contents

Eigenvalues and Eigenvectors

Definition. An **eigenvector** of a square matrix $A \in \mathbb{R}^{n \times n}$ with corresponding **eigenvalue** $\lambda \in \mathbb{C}$ is a vector $\mathbf{u} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ that gets mapped onto the scalar multiple $\lambda\mathbf{u}$ of itself by A:

$$A\mathbf{u} = \lambda\mathbf{u}$$

Eigenvectors are important in machine learning, because...

- Any symmetric matrix A (therefore any covariance matrix) can be decomposed as

$$A = U \Lambda U^\top$$

where U is an orthogonal matrix whose columns are the eigenvectors of A and where Λ is a diagonal matrix with the corresponding eigenvalues of A on its diagonal.

Eigenvectors are important in machine learning, because... (cont'd)

- Eigenvectors form the solution to an important class of optimization problems:

$$\max_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^\top A \boldsymbol{x} \qquad (A \text{ symmetric})$$

  The maximum is attained by the eigenvector $\mathbf{u}$ with the largest eigenvalue $\lambda_{\max}$.

$$\max_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^\top A \boldsymbol{x} = \mathbf{u}^\top A \mathbf{u} = \lambda_{\max} \mathbf{u}^\top \mathbf{u} = \lambda_{\max}$$

Computation of eigenvalues/eigenvectors

Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. Then for any eigenvector $\mathbf{u}$ it holds that

$$A\mathbf{u} = \lambda\mathbf{u} \quad \Leftrightarrow \quad (A - \lambda I)\mathbf{u} = 0$$

1. Compute the eigenvalues as roots of the **characteristic polynomial** $P(\lambda) := \det(A - \lambda I)$. This polynomial has degree n.
2. For all computed real eigenvalues λ_i determine a basis of the vector space $\{\mathbf{u}_i \in \mathbb{R}^n \mid (A - \lambda_i I)\mathbf{u}_i = 0\}$.

Eigenvalues/eigenvectors - useful facts

Let $A \in \mathbb{R}^{n \times n}$ be a square matrix. The following hold:

- There exist at most n real eigenvalues and at most n linearly independent eigenvectors.
- It is not guaranteed that there are n linearly independent eigenvectors even if the characteristic polynomial has n real roots.
- Eigenvectors with different eigenvalues are linearly independent.
- A has n pairwise different eigenvalues $\Rightarrow$ A has n linearly independent eigenvectors

Table of Contents

Diagonalizability

Definition. A square matrix $A \in \mathbb{R}^{n \times n}$ is **diagonalizable** if there exist an invertible matrix $S \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ s.t.

$$\Lambda = S^{-1} A S.$$

$A = S \Lambda S^{-1}$ is called the **eigendecomposition** of A.

The following hold:

- A has n linearly independent eigenvectors $\Leftrightarrow$ A is diagonalizable
- The columns of S are the eigenvectors of A, the diagonal of Λ contains the corresponding eigenvalues.

Properties of diagonalizable matrices

Let $A \in \mathbb{R}^{n\times n}$ be diagonalizable with $A = S\Lambda S^{-1}$.

- All eigenvalues nonzero $\Leftrightarrow$ A is invertible and

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

- Diagonalization makes it easier to exponentiate A:

$$A^p = S\Lambda^p S^{-1} \quad \text{for } p \in \mathbb{N}$$

- The determinant of A is the product of its eigenvalues.
- The trace of A is the sum of its eigenvalues.

Table of Contents

Eigenvectors and -values of symmetric matrices

For any symmetric matrix $A \in \mathbb{R}^{n \times n}$ the following hold:

- All eigenvalues of A are real.
- Eigenvectors with pairwise different eigenvalues are orthogonal.
- There always exist n orthogonal eigenvectors.
- If in addition A has full rank, it can be decomposed as

$$A = U \Lambda U^\top$$

  where U is an orthogonal matrix with eigenvectors of A as columns and Λ is a diagonal matrix with the corresponding eigenvalues of A on its diagonal.

Positive definiteness

Definition. A square matrix $A \in \mathbb{R}^{n \times n}$ is

$$\left.\begin{array}{ll} \textbf{positive definite} & \text{if } \boldsymbol{v}^\top A \boldsymbol{v} > 0 \\ \textbf{positive semi-definite} & \text{if } \boldsymbol{v}^\top A \boldsymbol{v} \geq 0 \\ \textbf{negative definite} & \text{if } \boldsymbol{v}^\top A \boldsymbol{v} < 0 \\ \textbf{negative semi-definite} & \text{if } \boldsymbol{v}^\top A \boldsymbol{v} \leq 0 \end{array}\right\} \text{ for all } \boldsymbol{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$$

Theorem. For any symmetric matrix A the following hold:

$$\begin{array}{lcl} A \text{ positive definite} & \Leftrightarrow & \text{all eigenvalues} > 0 \\ A \text{ positive semi-definite} & \Leftrightarrow & \text{all eigenvalues} \geq 0 \\ A \text{ negative definite} & \Leftrightarrow & \text{all eigenvalues} < 0 \\ A \text{ negative semi-definite} & \Leftrightarrow & \text{all eigenvalues} \leq 0 \end{array}$$

- Comprehensive collection of computation rules for matrices:
  K. B. Petersen, M. S. Pedersen (2007) **The Matrix Cookbook.**
  `http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3274`