# Recap: Linear Algebra

Stephanie Brandl

Machine Learning: TU Berlin

# Vector space

**Definition.** A *vector space* in $\mathbb{R}$ consists of a set $V$ equipped with two operations, $\dot{+}$ (addition) and $\cdot$ (scalar multiplication)

$$\dot{+} : V \times V \to V \qquad (\vartheta, \omega) \mapsto \vartheta \dot{+} \omega$$
$$\cdot : \mathbb{R} \times V \to V \qquad (\lambda, \vartheta) \mapsto \lambda \cdot \vartheta$$

- $(V, \dot{+})$ is an abelian group, satisfying:
  - i) $\forall \vartheta_1, \vartheta_2, \vartheta_3 \in V : \; \vartheta_1 \dot{+} (\vartheta_2 \dot{+} \vartheta_3) = (\vartheta_1 \dot{+} \vartheta_2) \dot{+} \vartheta_3$ (associativity)
  - ii) $\vartheta \dot{+} e = e \dot{+} \vartheta = \vartheta \;\; \forall \vartheta \in V$ (unique identity element $e \in V$)
  - iii) $\forall \vartheta \in V \; \exists! \; \vartheta^{-1} \in V : \quad \vartheta \dot{+} \vartheta^{-1} = \vartheta^{-1} \dot{+} \vartheta = e$
  - iv) $\forall \vartheta_1, \vartheta_2 \in V : \vartheta_1 \dot{+} \vartheta_2 = \vartheta_2 \dot{+} \vartheta_1$ (commutativity)
- $(V, \cdot)$ satisfies:
  - i) $\lambda(\mu\vartheta) = (\lambda\mu)\vartheta$ (associativity)
  - ii) $1\vartheta = \vartheta$ (unique identity element)
  - iii) $\lambda(\vartheta \dot{+} \omega) = \lambda\vartheta \dot{+} \lambda\omega$, $(\lambda \dot{+} \mu)\vartheta = \lambda\vartheta \dot{+} \mu\vartheta$ (distributivity)

# The Euclidean norm- length of a vector

**Definition.** The *(euclidean) norm* of a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{R}^d$ is given by

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{d} x_i^2}$$

For $\mathbf{x} \in \mathbb{R}^2$ , $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$

The distance of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ is then given by $\|\mathbf{x} - \mathbf{y}\|$.

# The scalar product

**Definition.** The *(standard) scalar product* of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ is given by:

$$\mathbf{x}^T\mathbf{y} = x_1y_1 + \ldots + x_dy_d.$$

Notice from the definition: $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}}$

What does the scalar product tell us?

- $\mathbf{x}^T\mathbf{y} = \cos(\angle\mathbf{x}, \mathbf{y}) \cdot \|\mathbf{x}\| \cdot \|\mathbf{y}\|$
- For $\|\mathbf{x}\| = 1$:
  $\mathbf{x}^T\mathbf{y}$ is the length of the orthogonal projection of $\mathbf{y}$ on $\mathbf{x}$

**Definition.** $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are *orthogonal* to each other if

$$\mathbf{x}^T\mathbf{y} = 0$$

# Matrices as linear transformations

**Definition.** A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **linear** if it satisfies

$$f(\lambda_1 \mathbf{x} + \lambda_2 \mathbf{y}) = \lambda_1 f(\mathbf{x}) + \lambda_2 f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda_1, \lambda_2 \in \mathbb{R}$

- if $f$ is linear, we have for each $\mathbf{x} \in \mathbb{R}^n$:

$$f(\mathbf{x}) = f(x_1 \mathbf{e}_1 + \ldots + x_n \mathbf{e}_n) = x_1 f(\mathbf{e}_1) + \ldots + x_n f(\mathbf{e}_n)$$

- $\Rightarrow$ The function is entirely determined by $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)$

# Matrices as linear transformations

$$f(\mathbf{x}) = f(x_1\mathbf{e}_1 + \ldots + x_n\mathbf{e}_n) = x_1 f(\mathbf{e}_1) + \ldots + x_n f(\mathbf{e}_n)$$

We can represent $f$ by a $m \times n$ matrix A

$$f(\mathbf{x}) = A \cdot \mathbf{x} = \begin{pmatrix} | & & | \\ f(\mathbf{e}_1) & \ldots & f(\mathbf{e}_n) \\ | & & | \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

The columns of $A$ are the images of the unit vectors.

# Basic matrix operations

- The **sum** $A + B$ of matrices $A, B$ is calculated entrywise
- The **scalar multiplication** $\lambda A$ of matrix $A$ and scalar $\lambda$ is calculated entrywise
- The **transpose** of matrix $A \in \mathbb{R}^{m \times n}$ is the matrix $A^T \in \mathbb{R}^{n \times m}$ formed by turning rows into columns
- For $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times k}$, the entries of the **matrix product** $AB \in \mathbb{R}^{m \times k}$ are given by the scalar product of the corresponding row of A and the corresponding column of B:

$$(AB)_{ij} = a_{i1}b_{1j} + \ldots + a_{in}b_{nj} = \mathbf{a}_{i\cdot} \cdot \mathbf{b}_{\cdot j}$$

# Rank

**Definition.** The *rank* of a matrix is the maximum number of linearly independent column vectors of A. (equivalent: rows).

**Definition.** A square matrix $A \in \mathbb{R}^{n \times n}$ is called *invertible*, if there exists a matrix $A^{-1}$ such that

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

A has full rank ( $\text{rang}(A) = n$ ) $\Leftrightarrow$ A is invertible.

**Useful properties:**

$$(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$$
$$(A^T)^{-1} = (A^{-1})^T$$
$$(A + B)^T = A^T + B^T$$
$$(\lambda \cdot A)^T = \lambda \cdot A^T$$
$$(A \cdot B)^T = B^T \cdot A^T$$

**Special matrices:**

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

- A is **orthogonal** if $AA^T = A^T A = I_n$.
  Orthogonale Matrices represent rotations or reflections.
- A is **symmetric** if $A = A^T$
- A is **diagonal** if all elements on the off-diagonal are $0$

# Eigenvectors

**Definition.** An *eigenvector* of a square matrix $A \in \mathbb{R}^{n \times n}$ with a corresponding *eigenvalue* $\lambda \in \mathbb{R}$ is a vector $\mathbf{v} \in \mathbb{R}^n \setminus \{0\}$ that, when multiplied by $A$ yields a scaled version of itself:

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

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

- $A$ has a maximum of $n$ eigenvalues and $n$ linear independent eigenvectors
- $A$ does not necessarily have $n$ linear independent eigenvectors

If $A$ is symmetric:

- The eigenvalues are real
- There are $n$ orthogonal eigenvectors
- It follows:

$$AU = U\Lambda$$

  where $U$ is an orthogonal matrix containing the eigenvectors in the columns and $\Lambda$ is a diagonal matrix with corresponding eigenvalues on the diagonal
- This yields the **eigendecomposition** of $A$

$$A = U\Lambda U^T$$