Groups, Fields & Euclidean Vector Spaces

Thomas Schnake, Naima Elosegui Borras, Tom Kaufmann

Machine Learning Group

May 21, 2024

Table of Contents

Organisation I

- This course serves as an **elective** for the modules *Machine Learning 1-X & 2* and *Cognitive Algorithms*.
- Course page: `https://isis.tu-berlin.de/user/index.php?id=38983`
- In order to pass the course, you need to pass a **test (90 minutes)** (which will determine your final grade). Requirement for taking the exam is reaching $\geq 50\%$ of the points on the homework sheets. Other than that, the number of points gained on the homework sheets does not affect your grade.

Organisation II

Homework:

- Please form groups of up to three students and submit the groups in the ISIS course page.
- Homework will be posted after each lecture and is due the following Tuesday at 10.00.
- Submit each homework in your groups as a single pdf file in the ISIS course page.
- You will need 50% of the points on the homework sheets to be allowed to take the exam.

The Exam will take place on July 2nd from 10:15 to 11:45

Organisation III

Weekly structure:

- Tuesdays 10.15 – 11.45: Lecture (Please submit homework until 10.00!)
- Thursdays 10.15 – 11.45: Discussion of exercise sheets

Class content:

- Week 1: Groups, fields, Euclidean vector spaces
- Week 2: Linear transformations, matrices & determinant
- Week 3: Differentiation & examples from ML
- Week 4: Probability theory
- Week 5: Selected Subject - Graph Convolutions in Deep Learning

Motivation for Today: Machine Learning = Learning from Data

Machine learning (e.g., image classification, spam detection, DNA segmentation, ...) is learning from data.

Data is usually represented as vectors, i.e. members of a vector space. This is true for one dimensional vectors (e.g., time series), two dimensional vectors (e.g., images), and higher dimensional vectors (e.g., text data).

General approach:

1. Extraction of features and representations in form of vectors
2. Application of methods from linear algebra and probability theory to learn a model from data
3. Application of the learned model to new data

Today's Content

Table of Contents

Set

Definition. A **set** is a collection of distinct objects. The objects in a set are called **elements** or **members**. **Examples:**

- $A = \{1, 2, 3\}$ means that the set A contains the elements 1, 2, and 3.
- Natural numbers: $\mathbb{N} = 1, 2, 3, \ldots$
- Integers: $\mathbb{Z} = \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$
- Rational numbers: $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}$
- Real numbers: $\mathbb{R} = \ldots, -1, \ldots, -0.5, \ldots, 0, \ldots, 1, \ldots, \pi, \ldots$

Basic concepts:

- **Membership**: $a \in A$ means a is an element of A.
- **Subset**: $A \subseteq B$ means every element of A is also an element of B.
- **Equivalence**: $A = B$ means $A \subseteq B$ and $B \subseteq A$.
- **Empty set**: The set with no elements is denoted by $\emptyset$.

Operation

Let S be a set. An **operation** $*\colon S \times S \to S$ is a mapping that maps two elements $a, b \in S$ onto another element $a * b \in S$.

$$*\colon S \times S \to S$$
$$(a,b) \mapsto a * b \qquad a, b, a * b \in S$$

Examples:

- Addition on integers: $a, b \in \mathbb{Z} \implies a + b \in \mathbb{Z}$
- Composition of rotations in 2D: $R_{\theta_1}, R_{\theta_2} \in SO(2) \implies R_{\theta_1} \cdot R_{\theta_2} \in SO(2)$

Counter-examples:

- Subtraction on natural numbers: $a, b \in \mathbb{N} \not\Longrightarrow a - b \in \mathbb{N}$ (e.g., $2 - 3 \notin \mathbb{N}$)
- Concatenation of strings in a given languages: $a, b \in$ English $\not\Longrightarrow a + b \in$ English (e.g., "hello" + "world" = "helloworld" is not an English word)

Group

Definition. A set $\mathcal{G}$ together with an operation $*\colon \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ is called a **group**, denoted $G = (\mathcal{G}, *)$, if all of the following hold:

1. $*$ is **associative**

$$\forall a, b, c \in \mathcal{G}\colon a * (b * c) = (a * b) * c,$$

2. there exists a unique **identity** or **neutral element** $e \in \mathcal{G}$ s.t.

$$\forall x \in \mathcal{G}\colon x * e = e * x = x$$

3. for every $a \in \mathcal{G}$ there exists a unique **inverse element** $a^{-1} \in \mathcal{G}$

$$\forall a \in \mathcal{G}\ \exists_1 a^{-1} \in \mathcal{G}\colon a * a^{-1} = a^{-1} * a = e$$

It is called **Abelian** or **commutative** if it also holds that

4. $*$ is commutative

$$\forall a, b \in \mathcal{G} : a * b = b * a$$

Group: Examples

Examples of abelian Groups:

- The set of integers $\mathbb{Z}$ with addition $(\mathbb{Z}, +)$.
  - Associative: $(a + b) + c = a + (b + c)$.
  - Identity: $0 \in \mathbb{Z}$, $a + 0 = 0 + a = a$.
  - Inverse: For $a \in \mathbb{Z}$, $-a \in \mathbb{Z}$, $a + (-a) = (-a) + a = 0$.
  - Commutative: $a + b = b + a$.
- The set of non-zero real numbers $\mathbb{R}^*$ with multiplication $(\mathbb{R}^*, \cdot)$.
  - Associative: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
  - Identity: $1 \in \mathbb{R}^*$, $a \cdot 1 = 1 \cdot a = a$.
  - Inverse: For $a \in \mathbb{R}^*$, $a^{-1} \in \mathbb{R}^*$, $a \cdot a^{-1} = a^{-1} \cdot a = 1$.
  - Commutative: $a \cdot b = b \cdot a$.

Group: Examples

Counter-Example:

- The set of natural numbers $\mathbb{N}$ with addition $(\mathbb{N}, +)$.
  - Lacks Inverse: For $a \in \mathbb{N}$, there is no $-a \in \mathbb{N}$.
- The set of integers $\mathbb{Z}$ with multiplication $(\mathbb{Z}, \cdot)$.
  - Lacks Inverse: For $a \in \mathbb{Z}$, there is no $a^{-1} \in \mathbb{Z}$.

Group that is not abelian:

- The set of 2×2 matrices with matrix multiplication.
  - Non-Commutative: In general, $AB \neq BA$ for matrices A and B.

Field

Definition. A set $\mathcal{F}$ together with two operations $\oplus\colon \mathcal{F} \times \mathcal{F} \to \mathcal{F}$ (addition) and $\odot\colon \mathcal{F} \times \mathcal{F} \to \mathcal{F}$ (multiplication) is called a **field**, denoted $F = (\mathcal{F}, \oplus, \odot)$, if all of the following hold:

1. $(\mathcal{F}, \oplus)$ forms an abelian group
2. $(\mathcal{F} \setminus \{0\}, \odot)$ forms an abelian group
3. distributive property
   $\forall a, b, c \in \mathcal{F}\colon a \odot (b \oplus c) = (a \odot b) \oplus (a \odot c)$

Examples: $(\mathbb{Q}, +, \cdot), (\mathbb{R}, +, \cdot), (\mathbb{C}, +, \cdot), (\mathbb{F}_2 = \{0, 1\}, +_{mod_2}, \cdot)$
Counter-example: $(\mathbb{Z}, +, \cdot)$, since $(\mathbb{Z} \setminus \{0\}, \cdot)$ is not an abelian group

Table of Contents

Vector space

Definition. Let $F = (\mathcal{F}, \oplus, \odot)$ be a field. A set $\mathcal{V}$ together with two operations

$$
\begin{aligned}
+ &\colon \mathcal{V} \times \mathcal{V} \to \mathcal{V} && (\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} + \boldsymbol{y} && \text{(vector addition)} \\
\cdot &\colon \mathcal{F} \times \mathcal{V} \to \mathcal{V} && (\lambda, \boldsymbol{x}) \mapsto \lambda \cdot \boldsymbol{x} && \text{(scalar multiplication)}
\end{aligned}
$$

is called an **F–vector space**, denoted $V = (\mathcal{V}, +, \cdot)$ if all of the following hold:

1. $(\mathcal{V}, +)$ forms an abelian group whose identity element is called the zero vector $\mathbf{0}$,
2. scalar and field multiplication are compatible
$$\forall \lambda, \mu \in \mathcal{F}\ \forall \boldsymbol{x} \in \mathcal{V} \colon \lambda \cdot (\mu \cdot \boldsymbol{x}) = (\lambda \odot \mu) \cdot \boldsymbol{x}$$
3. there exists a unique identity element $e \in \mathcal{F}$ s.t.
$$\forall \boldsymbol{x} \in \mathcal{V} \colon e \cdot \boldsymbol{x} = \boldsymbol{x}$$
4. scalar multiplication is distributive
$$\lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \boldsymbol{x} + \lambda \boldsymbol{y}, \quad (\lambda \oplus \mu) \cdot \boldsymbol{x} = \lambda \boldsymbol{x} + \mu \boldsymbol{x}$$

Vector Space: Examples

Examples:

- $\mathbb{R}^3$ over $\mathbb{R}$:
  - The set of all 3-dimensional vectors with real number scalars.
  - Vector addition is component-wise, and scalar multiplication is also component-wise.
- Polynomials of degree $\leq n$ over $\mathbb{R}$:
  - The set of all polynomials whose degree is less than or equal to n, with real number coefficients.
  - Vector addition is polynomial addition, and scalar multiplication is multiplication by a scalar.

Typical Fields and Vector Spaces in Machine Learning

Scalar Fields:
We assume the field $(\mathbb{R}, +, \cdot)$ for scalars. A scalar is a single number used to scale vectors in a vector space.

Vector Spaces:

Vector spaces are column vectors of the form $\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{R}^d = \mathcal{V}$

Vector addition:

$$\boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_d + y_d \end{bmatrix}$$

Scalar multiplication:

$$\lambda \boldsymbol{x} = \begin{bmatrix} \lambda x_1 \\ \vdots \\ \lambda x_d \end{bmatrix}$$

Vector subspace

Definition. Let V be an F-vector space and $\mathcal{U} \subseteq \mathcal{V}$. Then $U = (\mathcal{U}, +, \cdot)$ is called **vector subspace of V** if all of the following (closure) hold:

1. $\mathcal{U} \neq \varnothing$
2. $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{U} \Rightarrow \boldsymbol{x} + \boldsymbol{y} \in \mathcal{U}$
3. $\boldsymbol{x} \in \mathcal{U}, \lambda \in \mathcal{F} \Rightarrow \lambda \boldsymbol{x} \in \mathcal{U}$

If U is a vector subspace of V then U itself is a vector space.

Vector subspace: Examples

Example:

- Let $V = \mathbb{R}^3$ and $\mathbb{U}$ be the xy-plane, where $\mathbb{U} = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$. This satisfies closure under vector addition and scalar multiplication, making it a vector subspace of $\mathbb{R}^3$.

Counter-example:

- Consider $V = \mathbb{R}^3$ and $\mathbb{U} = \{(x, y) \mid x^2 + y^2 <= 1\}$. While $\mathbb{U}$ is not empty and closed under addition, it fails to satisfy closure under scalar multiplication for negative scalars, thus it is not a vector subspace of $\mathbb{R}^3$.

Linear combination

Definition. Let V be a F-vector space, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathcal{V}$, $n \in \mathbb{N}$.
A vector $\boldsymbol{x} \in \mathcal{V}$ is a **linear combination** of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ if there exist $\lambda_1, \ldots, \lambda_n \in \mathcal{F}$ s.t.

$$\boldsymbol{x} = \lambda_1 \boldsymbol{v}_1 + \ldots + \lambda_n \boldsymbol{v}_n = \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i.$$

Example:

- In $\mathbb{R}^2$, any vector (x, y) can be represented as a linear combination of the standard basis vectors $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$, where $\boldsymbol{x} = x\mathbf{i} + y\mathbf{j}$.
- In the vector space of polynomial functions of degree less than or equal to n, any polynomial function $f(x)$ can be represented as a linear combination of the basis polynomials $1, x, x^2, \ldots, x^n$, where $f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$.

Linear independence

Vectors are called linearly independent if none of them can be written as a linear combination of the others.

Definition. Let V be a F-vector space. Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathcal{V}$, $n \in \mathbb{N}$ are called **linearly independent** if

$$\lambda_1 \boldsymbol{v}_1 + \ldots + \lambda_n \boldsymbol{v}_n = 0 \qquad (\lambda_1, \ldots, \lambda_n \in \mathcal{F})$$

implies that

$$\lambda_1 = \ldots = \lambda_n = 0.$$

Equation above is the same as:

$$-\frac{1}{\lambda_n}(\lambda_1 \boldsymbol{v}_1 + \ldots) = \boldsymbol{v}_n \qquad (\lambda_n \neq 0)$$

$$\lambda_1^* \boldsymbol{v}_1 + \ldots = \boldsymbol{v}_n \qquad (\lambda_1^*, \ldots, \lambda_n^* \in \mathcal{F})$$

Span

Definition. Let V be a F-vector space and $\mathcal{U} = \{\boldsymbol{u}_i \mid i = 1, \ldots, m\} \subseteq \mathcal{V}$ an arbitrary subset. The **span** of $\mathcal{U}$, denoted by

$$\operatorname{span}_F(\mathcal{U}),$$

is the set of all $\boldsymbol{x} \in \mathcal{V}$ that are linear combinations of finite subsets of $\mathcal{U}$. That is for all $\boldsymbol{x} \in \operatorname{span}_F(\mathcal{U})$ there exists $r \in \mathbb{N}$, indices $i_1, \ldots i_r \in \{1, \ldots, m\}$ and scalars $\lambda_1, \ldots, \lambda_r \in \mathcal{F}$ s.t.

$$\boldsymbol{x} = \lambda_1 \boldsymbol{u}_{i_1} + \ldots + \lambda_r \boldsymbol{u}_{i_r}.$$

$\mathcal{W} = \operatorname{span}_F(\mathcal{U}) \subset \mathcal{V}$ induces a vector subspace $(\mathcal{W}, +, \cdot)$ of V.

Geometric interpretation

Grant Sanderson's great visualizations of span: `https://youtu.be/k7RM-ot2NWY`

Spanning set & basis

Definition. Let V be a F-vector space and $\mathcal{M} \subseteq \mathcal{V}$.

1. $\mathcal{M}$ is a **spanning set** of $\mathcal{V}$ if $\operatorname{span}(\mathcal{M}) = \mathcal{V}$.
2. $\mathcal{M}$ is a **basis** of $\mathcal{V}$ if it is a spanning set of $\mathcal{V}$ and the vectors in $\mathcal{M}$ are linearly independent.
3. $\mathcal{V}$ is **finitely generated** if it has a finite spanning set.

Spanning set & basis: Examples

1. **Spanning Set**:
   - *Example*: $\mathcal{M} = \{(1,0),(1,1),(3,3),(-3,1)\}$ in $\mathbb{R}^2$. This set spans $\mathbb{R}^2$.
   - *Counterexample*: $\mathcal{M} = \{(1,0)\}$ in $\mathbb{R}^2$. This set does not span $\mathbb{R}^2$.
2. **Basis**:
   - *Example*: $\mathcal{M} = \{(1,0),(0,1)\}$ in $\mathbb{R}^2$. This set forms a basis for $\mathbb{R}^2$ because it spans $\mathbb{R}^2$ and its vectors are linearly independent.
   - *Counterexample*: Let $\mathcal{M} = \{(1,0),(0,1),(1,1)\}$ in $\mathbb{R}^2$. Although it does span $\mathbb{R}^2$, its vectors are linearly dependent.
3. **Finitely Generated**:
   - *Example*: $\mathbb{R}^2$ is finitely generated because it can be spanned by a finite set, such as $\{(1,0),(0,1)\}$.
   - *Counterexample*: The vector space of all polynomials is not finitely generated.

Basis theorem

Theorem and Definition. Any two bases of a finitely generated F-vector space V have the same size (number of vectors). This number is called the **dimension** of the vector space V.

Examples:

- $\mathbb{R}^3$: The vector space of 3-dimensional real vectors. It has a basis consisting of three linearly independent vectors, such as $\{(1,0,0),(0,1,0),(0,0,1)\}$. Therefore, the dimension of $\mathbb{R}^3$ is 3.
- $F^{2\times 2}$: The vector space of 2×2 matrices over a field F. It has a basis consisting of four linearly independent matrices. Therefore, the dimension of $F^{2\times 2}$ is 4.

Table of Contents

Scalar product

Definition. Let V be a real vector space ($\mathbb{R}$-vector space). A **scalar product** or **inner product on V** is a mapping $\langle \cdot\,, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ with the following properties for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{V}$ and $\lambda \in \mathbb{R}$:

1. bilinearity:

$$\begin{aligned} \langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle &= \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \lambda \boldsymbol{y} \rangle \\ \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle &= \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle \\ \langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle &= \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle \end{aligned}$$

2. symmetry: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$
3. positive definiteness: $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ and $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$

Scalar product: examples

- Standard scalar product on $\mathbb{R}^n$:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := x_1 y_1 + \ldots + x_n y_n = \boldsymbol{x}^\top \boldsymbol{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta \qquad (\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n)$$

- On $\mathbb{R}^n$ with $\lambda_1, \ldots, \lambda_n \geq 0$:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \lambda_1 x_1 y_1 + \ldots + \lambda_n x_n y_n \qquad (\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n)$$

- On $\mathcal{F} = L_2(\mathcal{N}) = \left\{ X \colon \mathcal{N} \to \mathbb{R} \mid \int_{\mathcal{N}} X(n)^2 dn < \infty \right\}$, the space of square-integrable functions on a compact set $\mathcal{N} \subset \mathbb{R}^n$

$$\langle X, Y \rangle := \int_{\mathcal{N}} X(n) Y(n) dn \qquad (X, Y \in \mathcal{F})$$

Norm – length of a vector

Definition. Let V be a real vector space. A **norm** is a mapping $\|\cdot\|\colon \mathcal{V} \to \mathbb{R}^+$ with the following properties $\forall \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V},\ \forall \lambda \in \mathbb{R}$:

1. definiteness: $\|\boldsymbol{v}\| \geq 0$ and $\|\boldsymbol{v}\| = 0 \Leftrightarrow \boldsymbol{v} = \mathbf{0}$
2. triangle inequality: $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$
3. homogeneity: $\|\lambda \boldsymbol{v}\| = |\lambda| \|\boldsymbol{v}\|$

Examples on $\mathcal{V} = \mathbb{R}^n$:

- 2-norm: $\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
- 1-norm: $\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$
- max-norm: $\|\boldsymbol{x}\|_\infty = \max_{i=1}^n |x_i|$

Euclidean vector spaces

Definition. A real vector space V of finite dimension, together with a scalar product $(V, \langle \cdot\,, \cdot \rangle)$ is called **Euclidean vector space**.

In a Euclidean vector space the **Euclidean norm** is induced by the scalar product:

$$\begin{aligned} \| \cdot \| : \mathcal{V} &\to \mathbb{R}^+ \\ \boldsymbol{v} &\mapsto \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \end{aligned}$$

Cauchy-Schwarz Inequality und Triangle Inequality

Theorem. Let $(V, \langle \cdot\, , \cdot \rangle)$ be a Euclidean vector space. Then the **Cauchy-Schwarz inequality** holds:

$$\forall \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V} \colon |\langle \boldsymbol{v}, \boldsymbol{w} \rangle| \leq \|\boldsymbol{v}\| \|\boldsymbol{w}\|.$$

Note. Naturally, the norm induced by the scalar product on V satisfies the **triangle inequality**:

$$\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|.$$

Orthogonal vectors

Definition. Two vectors $\boldsymbol{v}, \boldsymbol{w}$ of a Euclidean vector space are **orthogonal** ($\boldsymbol{v} \perp \boldsymbol{w}$) if $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = 0$.

Definition. Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in a Euclidean vector space are **orthonormal** or form an **orthonormal system** if $\|\boldsymbol{v}_i\| = 1$ for all i and $\boldsymbol{v}_i \perp \boldsymbol{v}_j$ for $i \neq j$.

Orthonormal basis

Definition. A basis of orthonormal vectors in an Euclidean vector space is called **orthonormal basis**.

Theorem. If $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ form an orthonormal basis of a Euclidean vector space $(V, \langle \cdot\,, \cdot \rangle)$, then for all $\boldsymbol{x} \in \mathcal{V}$:

$$\boldsymbol{x} = \sum_{i=1}^{n} \langle \boldsymbol{x}, \boldsymbol{u}_i \rangle \, \boldsymbol{u}_i$$

The Gram-Schmidt process can transform any basis into an orthonormal basis for the same vector space.

Summary

- Introduction to Sets, Groups, and Fields
- Explanation of Vector Spaces, Subspaces, and Linear Independence
- Overview of Euclidean Vector Spaces including scalar product, Norm, and Orthogonality