MADS-MMS – Mathematics and Multivariate Statistics

Analytic Geometry

Prof. Dr. Stephan Doerfel

FACHHOCHSCHULE KIEL
University of Applied Sciences

SCAN ME

Moodle (SoSe 2025)

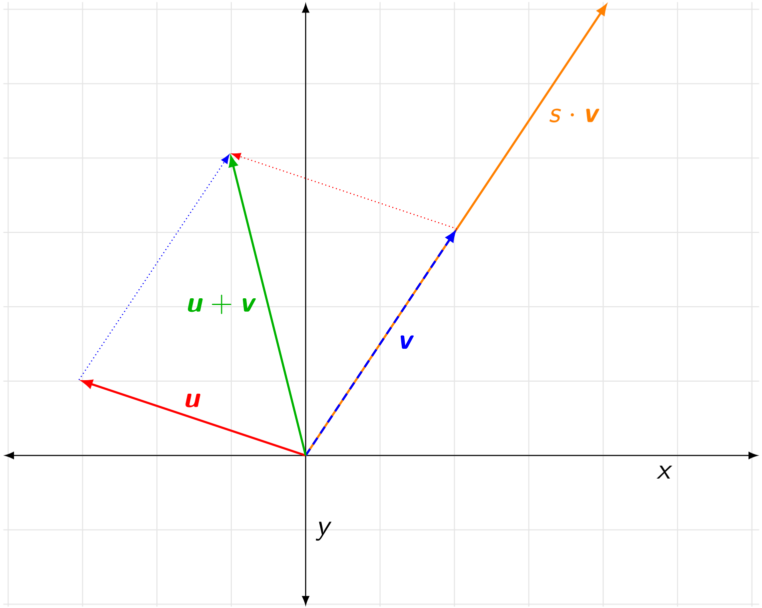
Agenda

Outline

Motivation

- one heavily exploited feature in machine learning are geometric relationships between instances
- kNN exploits distances between instances
- SVMs measure distances between planes and instances
- k-means groups instances with short distances between each other
- text-mining uses the angle between long vectors as similarity measure

Example 1 – $\mathbb{R}^2$ – Geometric Interpretation



Example 1 – $\mathbb{R}^2$ – Geometric Interpretation

- the geometric interpretation works similarly in $\mathbb{R}^n$ with $n > 2$
- it allows computing various geometric entities, like planes, volumes, distances, angles, ...

Chapter Goals

- mathematical foundations of geometry
- inner products (e.g. for distance functions, for SVMs)
- distance functions (e.g. for clustering)
- geometric figures (e.g. for SVMs)

Outline

Inner Product

Definition 1 (Inner Product)

Let V be a real-valued vector space. Then a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called an **inner product**, if for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $r, s \in \mathbb{R}$ holds

1. $\langle r\boldsymbol{u} + s\boldsymbol{v}, \boldsymbol{w} \rangle = r\langle \boldsymbol{u}, \boldsymbol{w} \rangle + s\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ and $\langle \boldsymbol{u}, r\boldsymbol{v} + s\boldsymbol{w} \rangle = r\langle \boldsymbol{u}, \boldsymbol{v} \rangle + s\langle \boldsymbol{u}, \boldsymbol{w} \rangle$ (**bilinear**)
2. $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$ (**symmetric**)
3. $\langle \boldsymbol{v}, \boldsymbol{v} \rangle > 0$ for $\boldsymbol{v} \neq 0$ and $\langle 0, 0 \rangle = 0$ (**positive definite**)

Example in $\mathbb{R}^2$:

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = u_1 v_1 - (u_1 v_2 + u_2 v_1) + 2 u_2 v_2$$

The Dot Product

- In the vector space $\mathbb{R}^n$, the dot product or scalar product

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle := \sum_{i=1}^{n} u_i v_i$$

  is an inner product.
- $\mathbb{R}^n$ together with the dot product is called a **Euclidean vector space**.

Outline

Vector Norm

Definition 2

A **norm** on a vector space is a function $\|\cdot\| : V \to \mathbb{R}$, such that for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and $r \in \mathbb{R}$ holds

1. $\|r\boldsymbol{v}\| = |r| \cdot \|v\|$
2. $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$ (**triangle inequality**)
3. $\|\boldsymbol{v}\| \geq 0$ and $\|\boldsymbol{v}\| = 0 \iff \boldsymbol{v} = 0$

The norm of a vector can be understood as its **length**.

Example: Manhattan Norm

$$\|\boldsymbol{v}\|_1 = \sum_{i=1}^{n} |v_i|$$

- sum the absolute values of a vector
- use case: travel distance in manhattan (taxicab norm)
- use case: regularization (in machine learning)
- What does a circle look like?

Example: Euclidean Norm

$$\|\boldsymbol{v}\|_2 := \sqrt{\sum_{i=1}^{n} v_i^2}$$

- square root of the sum of squares
- corresponds to the "intuitive" length of vector
- the length from Euclidean geometry
- What does a circle look like?

Observation:

$$\|\boldsymbol{v}\|_2 = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

Induced Norm

- If $\langle \cdot, \cdot \rangle$ is the inner product of a vector space, then it induces a norm on the vector space: $\|\boldsymbol{v}\| := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$.
- The dot product induces the Euclidean norm.
- Not each norm is an induced norm (e.g. Manhattan norm).

Unit Vectors

Given a vector $\boldsymbol{v}$. How do we get a vector with the same orientation, but with length 1?

$$\frac{1}{\|\boldsymbol{v}\|}\boldsymbol{v}$$

Proof: $\|\frac{1}{\|v\|}\boldsymbol{v}\| = \frac{1}{\|v\|}\|\boldsymbol{v}\| = 1$.

Distance Function

Definition 3

Let O be a set of objects. A **distance function** is a function $d : O \times O \to \mathbb{R}_{\geq 0}$ such that

1. $d(o_1, o_2) = d(o_2, o_1)$
2. $d(o_1, o_2) = 0 \iff o_1 = o_2$

Definition 4

A distance function d is called a metric, if for all o_1, o_2, o_3 the triangle inequality holds:

$$d(o_1, o_3) \leq d(o_1, o_2) + d(o_2, o_3)$$

The general definition of distance allows for various inclusions and combinations of an objects features.

Induced Metric

- In a vector space V with inner product $\langle \cdot, \cdot \rangle$,

$$d : (\boldsymbol{u}, \boldsymbol{v}) \mapsto \sqrt{\langle \boldsymbol{u} - \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v} \rangle}$$

  is a distance metric.
- In a vector space V, a norm induces a distance metric by

$$d : (\boldsymbol{u}, \boldsymbol{v}) \mapsto \|\boldsymbol{u} - \boldsymbol{v}\|.$$

The different inner products and their distance function properties are the key ingredient in SVMs using the kernel trick.

Distance functions: Minkowski-Metrics

Definition 5 (Minkowski-Metric)

Let $\boldsymbol{u} = (u_1, \ldots, u_n)$ and $\boldsymbol{v} = (v_1, \ldots, v_n)$ be real-valued vectors and $p \in \mathbb{R}_{>0}$. For $p \geq 1$, the Minkowski-Metric (L_p-metric) is defined as:

$$d_p(\boldsymbol{u}, \boldsymbol{v}) := \sqrt[p]{\sum_{i=1}^{n} |u_i - v_i|^p}$$

Often used Minkowski-Metrics:

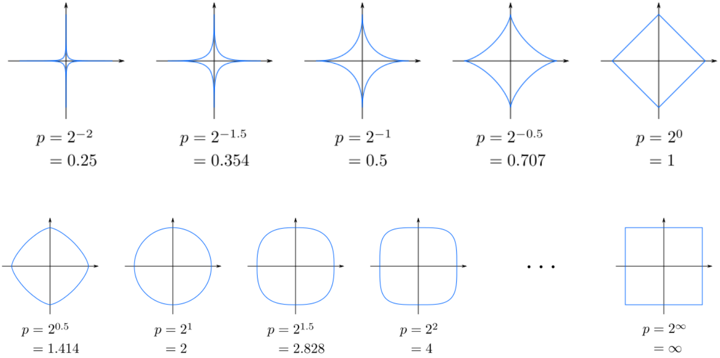
- Manhattan ($p = 1$): $d_1(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{n} |u_i - v_i|$
- Euclidean ($p = 2$): $d_2(\boldsymbol{u}, \boldsymbol{v}) = \sqrt[2]{\sum_{i=1}^{n} (u_i - v_i)^2}$
- Maximum ($p = \infty$): $d_\infty(\boldsymbol{u}, \boldsymbol{v}) = \max_{i=1}^{n} |u_i - v_i|$

Weighted Minkowski-Metric

- Sometimes we want to emphasize particularly important dimensions.
- The Minkowski-Metrics can be extended using weights ω_k:

$$d_p^{\omega}(\boldsymbol{u}, \boldsymbol{v}) = \sqrt[p]{\sum_{i=1}^{n} \omega_i |u_i - v_i|^p}$$

Unit Circles in Different Minkowski-Metrics



Source: https://commons.wikimedia.org/wiki/User:Waldir

Other Distance Functions

- For categorical attributes, the Hamming-Distance:

$$d(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{n} \delta(u_i, v_i) \quad \text{with} \quad \delta(u_i, v_i) = \begin{cases} 0 & \text{if } u_i = v_i \\ 1 & \text{otherwise} \end{cases}$$

- For sets A and B, the Jaccard Distance:

$$J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

- For strings the Levenshtein distance and the Damerau–Levenshtein distance between two strings measure the number of edit operations for turning one string into the other.
- many more, see e.g. implementations of `sklearn.neighbors.DistanceMetric`

Similarity Function

- Distance functions measure the opposite of similarity. Higher means more distance, thus less similarity.
- When we talk about measuring similarity, we usually rather mean distance.

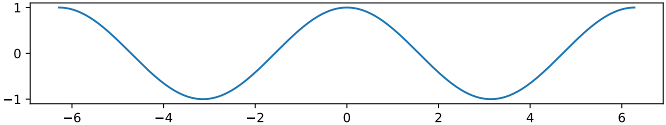
Exercises 1–3

Outline

The Cosine of Two Vectors

- Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$, then for two vectors $\boldsymbol{u}, \boldsymbol{v} \neq 0$
$$\cos \alpha = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\|}$$
is the cosine of the angle between the two vectors
- Hereby the norm is induced by the inner product
- With the dot product (thus Euclidean norm), the angles correspond to our usual intuition of angles

Recap Cosine



- cos is a trigonometric function
- $\cos(x + 2k\pi) = \cos(x)$ for $k \in \mathbb{Z}$ (periodic)
- $\cos(0) = 1$ (0°)
- $\cos(\frac{\pi}{2}) = 0$ (90°)
- $\cos(\pi) = -1$ (180°)

Orthogonal Vectors

Two vectors are orthogonal (angle 90°) if their inner product is zero.

Cosine Similarity

$$d(\boldsymbol{u}, \boldsymbol{v}) = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\|}$$

- an actual similarity function (higher means more similar)
- ignores the length of vectors, only orientation is relevant
- 1 means identical orientation
- -1 means inverse orientation
- $1 - d(\boldsymbol{u}, \boldsymbol{v})$ is often used as distance measure, however not a proper metric (triangle inequality)

Example Cosine Similarity for Texts

Representing text documents as vectors:

- Simple Version: dimensions are words, each entry is the frequency of the term in the document
- Tf-idf: dimensions are words, each entry is the product of its frequency in the document (tf) and the negative log of the share of documents that contain the term (idf).
  - Advantage: Reduce the influence of very common words, emphasize document-specific words.

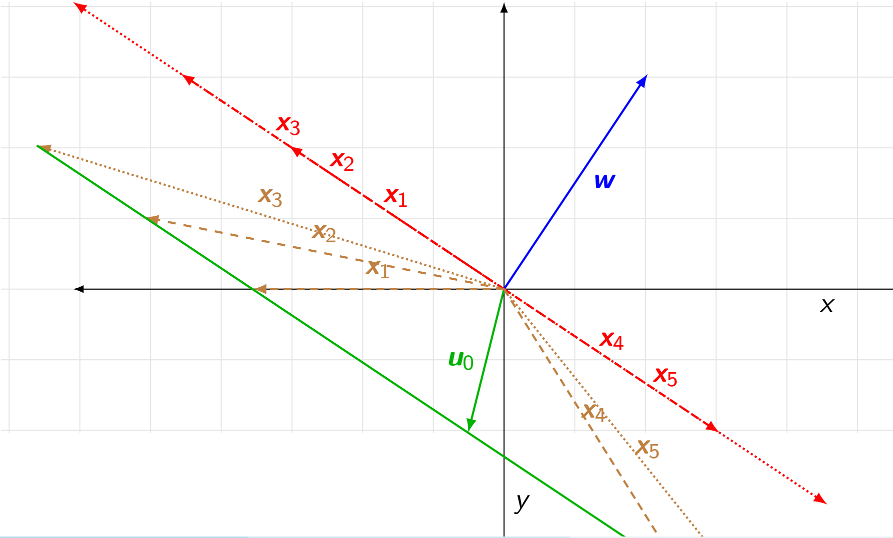
In both cases: very long, sparse vectors.

- cosine similarity works well with texts of different length
- a text and the same text appended to itself have cosine distance 0 but would have high values with many other distance functions (e.g. Euclidean)

Outline

A Line in $\mathbb{R}^2$

Let's consider $\mathbb{R}^2$. How can we describe a line?



A Line in $\mathbb{R}^2$

- all vectors that are orthogonal to a given vector $\boldsymbol{w} \neq 0$ form a line through the origin in $\mathbb{R}^2$
- adding one vector $\boldsymbol{u}_0$ moves the line away from the origin
- thus, $\boldsymbol{x}$ is on the line, if and only if $\boldsymbol{x} - \boldsymbol{u}_0$ is orthogonal to $\boldsymbol{w}$, meaning

$$\langle \boldsymbol{x} - \boldsymbol{u}_0, \boldsymbol{w} \rangle = 0$$

- equivalent:

$$\langle \boldsymbol{x}, \boldsymbol{w} \rangle - \langle \boldsymbol{u}_0, \boldsymbol{w} \rangle = 0$$

- In this model: $\boldsymbol{w}$ and u_0 are fix parameters that determine the line.

Alternative Line Descriptions in $\mathbb{R}^2$

$$\langle \boldsymbol{x} - \boldsymbol{u}_0, \boldsymbol{w} \rangle = 0 \quad \text{with } \boldsymbol{w}, u_0 \in \mathbb{R}^2$$

$$\langle \boldsymbol{x}, \boldsymbol{w} \rangle + b = 0 \quad \text{with } \boldsymbol{w} \in \mathbb{R}^2, b \in \mathbb{R}$$

$$x_2 = m x_1 + n \quad \text{with } \boldsymbol{x} = (x_1, x_2), m, n \in \mathbb{R}$$

With $\langle \cdot, \cdot \rangle$ being the dot product, the first two are equivalent. The third is the linear function description of a line. It is however not fully equivalent to the other two.

Exercises Homework: Prove 1 $\iff$ 2, Here: Prove 2 $\iff$ 3.

Hyperplane

Definition 6 (Hyperplane)

A **Hyperplane** in an n-dimensional vector space $\mathbb{R}^n$ is an $(n-1)$-dimensional affine subspace.

- n=1 → point
- n=2 → line
- n=3 → (regular) plane

The construction of hyperplanes is exactly analogous to the hyperplanes in $\mathbb{R}^2$ (lines).

Distance between Point and Hyperplane

Given a hyperplane in $\mathbb{R}^n$ as $\langle \boldsymbol{x}, \boldsymbol{w} \rangle + b = 0$ and a vector $\boldsymbol{v} \in \mathbb{R}^n$. Assuming $\boldsymbol{w} \neq 0$, what is the distance between $\boldsymbol{v}$ and the hyperplane?

- to $\boldsymbol{v}$, add a vector $\boldsymbol{z}$ that is orthogonal to the hyperplane, such that the sum is on the hyperplane:

$$\langle \boldsymbol{v} + \boldsymbol{z}, \boldsymbol{w} \rangle + b = 0 \text{ and } \boldsymbol{z} = r\boldsymbol{w} \text{ with } r \in \mathbb{R}$$

- Thus:

$$\langle \boldsymbol{v} + r\boldsymbol{w}, \boldsymbol{w} \rangle + b = 0 \Longrightarrow |r| = \frac{-(b + \langle \boldsymbol{v}, \boldsymbol{w} \rangle)}{\langle \boldsymbol{w}, \boldsymbol{w} \rangle}$$

- The distance we are looking for is:

$$\|\boldsymbol{z}\| = \|r\boldsymbol{w}\| = |r| \, \|\boldsymbol{w}\| = \frac{-(b + \langle \boldsymbol{v}, \boldsymbol{w} \rangle)}{\|\boldsymbol{w}\|^2} \|\boldsymbol{w}\| = \frac{-(b + \langle \boldsymbol{v}, \boldsymbol{w} \rangle)}{\|\boldsymbol{w}\|}$$

Exercises 4–6