Session 4: Norms and inner-products

Optimization and Computational Linear Algebra for Data Science

Léo Miolane

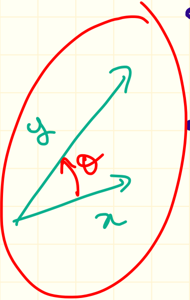
Contents

Norms and inner-products

Questions

Norms in machine learning:

- $\|\cdot\|_2$ Euclidean norm $\rightarrow$ measure distances.



- $x \cdot y$ Euclidean dot product

$$\cos(\theta) = \frac{x \cdot y}{\|x\|_2 \|y\|_2} \in [-1, 1]$$

- Use norms for "regularization"

minimize $\text{Loss}(\text{data}, x) \begin{cases} + \lambda \|x\|_2^2 \\ + \lambda \|x\|_1 \end{cases}$ with $x \in \mathbb{R}^n$

Questions

Questions

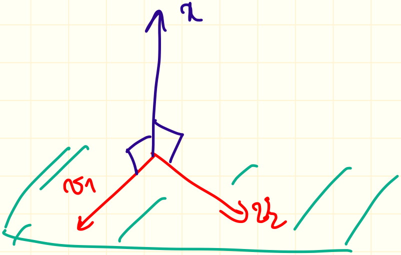
Orthogonality

Definition

Definition

- We say that vectors x and y are *orthogonal* if $\langle x, y\rangle = 0$. We write then $x \perp y$.
- We say that a vector x is orthogonal to a set of vectors A if x is orthogonal to all the vectors in A. We write then $x \perp A$.

Exercise: If x is orthogonal to $v_1, \ldots, v_k$ then x is orthogonal to any linear combination of these vectors i.e. $x \perp \mathrm{Span}(v_1, \ldots, v_k)$.



if $y \in \mathrm{Span}(v_1 \ldots v_k)$

then $y = \alpha_1 v_1 + \cdots + \alpha_k v_k$

$\langle x, y\rangle = \alpha_1 \underbrace{\langle x, v_1\rangle}_{=0} + \cdots + \alpha_k \underbrace{\langle x, v_k\rangle}_{=0}$

$= 0$

Pythagorean Theorem

Theorem (Pythagorean theorem)

Let $\|\cdot\|$ be the norm induced by $\langle\cdot,\cdot\rangle$. For all $x, y \in V$ we have

$$x \perp y \iff \|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

Proof.

$$\|x+y\|^2 = \langle x+y, x+y\rangle$$

$$= \|x\|^2 + 2\langle x,y\rangle + \|y\|^2$$

□

Application to random variables

$V = \{$ random variables with finite second moment $\}$

For X, Y, we define $\boxed{\langle X, Y\rangle = \mathbb{E}[XY]}$

the norm induced is $\|X\| = \sqrt{\langle X, X\rangle} = \sqrt{\mathbb{E}X^2}$

- Let's assume that X, Y have zero mean.

$X \perp Y \iff \mathbb{E}[XY] = 0 \iff \mathrm{Cov}(X, Y) = 0$

By Pyth. thm. this is equivalent to

$$\underbrace{\|X+Y\|^2}_{\mathbb{E}[(X+Y)^2]} = \|X\|^2 + \|Y\|^2$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$\cancel{\mathbb{E}[X^2] = +\infty}$

Orthogonal & orthonormal families

Definition

We say that a family of vectors $(v_1, \dots, v_k)$ is:

- *orthogonal* if the vectors $v_1, \dots, v_n$ are pairwise orthogonal, i.e. $\langle v_i, v_j \rangle = 0$ for all $i \neq j$.
- *orthonormal* if it is orthogonal and if all the v_i have unit norm: $\|v_1\| = \cdots = \|v_k\| = 1$.

Example:
- the canonical basis of $\mathbb{R}^n$ is orthonormal
- $\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right)$ is orthogonal but not orthonormal

Coordinates in an orthonormal basis

Proposition

A vector space of finite dimension admits an orthonormal basis.

Proposition

Assume that $\dim(V) = n$ and let $(v_1, \dots, v_n)$ be an **orthonormal** basis of V. Then the coordinates of a vector $x \in V$ in the basis $(v_1, \dots, v_n)$ are $(\langle v_1, x\rangle, \dots, \langle v_n, x\rangle)$:

$$x = \langle v_1, x\rangle v_1 + \cdots + \langle v_n, x\rangle v_n.$$

Proof: we have $x = \alpha_1 v_1 + \dots + \alpha_n v_n$ for some $\alpha_1 \dots \alpha_n \in \mathbb{R}$

$$\langle x, v_i \rangle = \langle \alpha_1 v_1 + \dots + \alpha_n v_n, v_i \rangle = \alpha_i \underbrace{\|v_i\|^2}_{=1}$$

Coordinates in an orthonormal basis

Let $x, y \in V = \mathbb{R}^n$

$$x = \underbrace{\langle v_1, x\rangle}_{\alpha_1} v_1 + \dots + \underbrace{\langle v_n, x\rangle}_{\alpha_n} v_n$$

$$y = \underbrace{\langle v_1, y\rangle}_{\beta_1} v_1 + \dots + \underbrace{\langle v_n, y\rangle}_{\beta_n} v_n$$

$$\langle x, y\rangle = \langle \alpha_1 v_1 + \dots + \alpha_n v_n, \beta_1 v_1 + \dots + \beta_n v_n \rangle$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j \underbrace{\langle v_i, v_j \rangle}_{} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$= \alpha_1 \beta_1 + \dots + \alpha_n \beta_n$$

$$\|x\| = \sqrt{\alpha_1^2 + \dots \alpha_n^2}$$

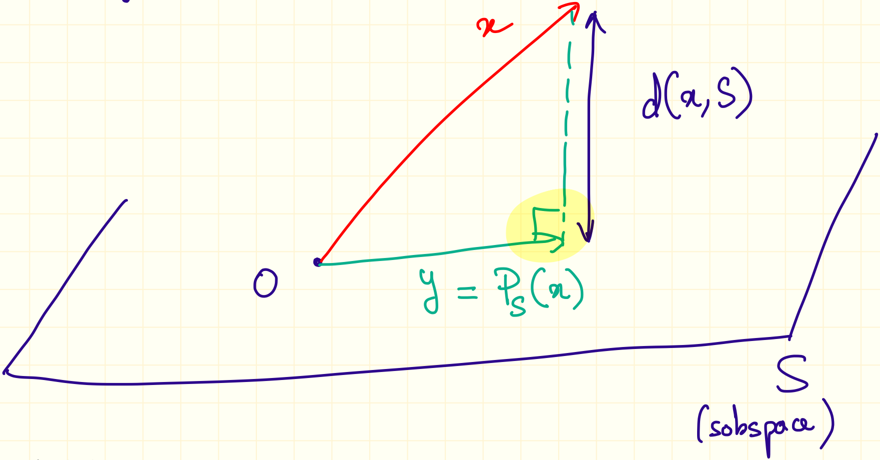
Proof

Orthogonal projection

Picture

From now, $\langle \cdot, \cdot \rangle$ denotes the Euclidean dot product, and $\| \cdot \|$ the Euclidean norm.

What is the vector y of S which is the closest to x?



Orthogonal projection and distance

Definition

Let S be a subspace of $\mathbb{R}^n$. The *orthogonal projection* of a vector x onto S is defined as the vector $P_S(x)$ in S that minimizes the distance to x:

$$P_S(x) \stackrel{\text{def}}{=} \underset{y \in S}{\arg\min} \|x - y\|.$$

the vector $y \in S$ that minimizes $\|x-y\|$

The distance of x to the subspace S is then defined as

$$d(x, S) \stackrel{\text{def}}{=} \underbrace{\min_{y \in S} \|x - y\|} = \underbrace{\|x - P_S(x)\|}.$$

- if $x \notin S$ then $x \neq P_S(x)$
- if $x \in S$ then $x = P_S(x)$

Computing orthogonal projections

Proposition

Let S be a subspace of $\mathbb{R}^n$ and let $(v_1, \dots, v_k)$ be an **orthonormal basis** of S. Then for all $x \in \mathbb{R}^n$,

$$P_S(x) = \underbrace{\langle v_1, x\rangle v_1 + \cdots + \langle v_k, x\rangle v_k}.$$

Proof: Let $y \in S$, $y = \alpha_1 v_1 + \dots + \alpha_k v_k$ for some $\alpha_1 \dots \alpha_k \in \mathbb{R}$

$$\|x-y\|^2 = \|x\|^2 \overbrace{- 2\langle x, y\rangle + \|y\|^2}$$

- $\|y\|^2 = \sum_{i=1}^{k} \alpha_i^2$
- $\langle x, y\rangle = \langle x, \alpha_1 v_1 + \dots + \alpha_k v_k\rangle = \sum_{i=1}^{k} \alpha_i \langle x, v_i\rangle$

Proof

$$\|x-y\|^2 = \|x\|^2 + \sum_{i=1}^{k} \alpha_i^2 - 2\alpha_i \langle x, v_i\rangle$$

minimize this wrt α_i

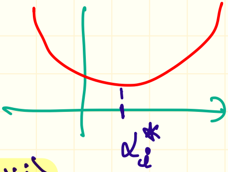
The α_i minimizing is $\boxed{\alpha_i = \langle x, v_i\rangle}$

Conclusion: the minimizer y is given by

$$P_S(x) = \langle x, v_1\rangle v_1 + \dots + \langle x, v_k\rangle v_k$$

$$f(\alpha_i) = \alpha_i^2 - 2\alpha_i \langle x, v_i\rangle$$

$$f'(\alpha_i^*) = 0 \quad \Big| \quad f'(\alpha_i) = 2\alpha_i - 2\langle x, v_i\rangle = 0 \rightarrow \alpha_i^* = \langle x, v_i\rangle$$



Consequence

$$P_S(x) = \langle x, v_1\rangle v_1 + \dots + \langle x, v_k\rangle v_k$$

Let define $V = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{pmatrix}$

- P_S is linear
- VV^T is the matrix of P_S

Claim : $\boxed{P_S(x) = VV^T x}$

Why? $\rightarrow$

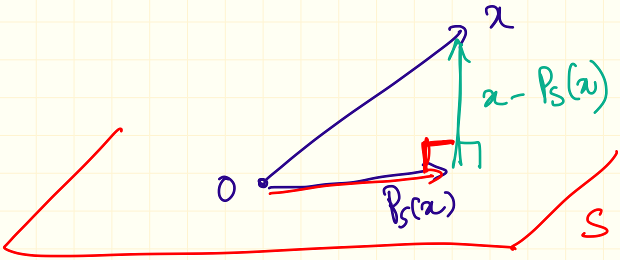
$$V^T x = \begin{pmatrix} — & v_1^T & — \\ & \vdots & \\ — & v_k^T & — \end{pmatrix} \begin{pmatrix} \\ x \\ \\ \end{pmatrix} = \begin{pmatrix} \langle v_1, x\rangle \\ \vdots \\ \langle v_k, x\rangle \end{pmatrix}$$

$$VV^T x = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_k \\ | & & | \end{pmatrix} \begin{pmatrix} \langle v_1, x\rangle \\ \vdots \\ \langle v_k, x\rangle \end{pmatrix} = \langle v_1, x\rangle v_1 + \dots + \langle v_k, x\rangle v_k.$$

Corollary

For all $x \in \mathbb{R}^n$,

- $x - P_S(x)$ is orthogonal to S.
- $\|P_S(x)\| \leq \|x\|$.



Proof of Cauchy-Schwarz inequality

Cauchy-Schwarz inequality

Theorem

Let $\|\cdot\|$ be the norm induced by the inner product $\langle\cdot,\cdot\rangle$ on the vector space V. Then for all $x, y \in V$:

$$\boxed{|\langle x, y\rangle| \leq \|x\| \, \|y\|.} \tag{1}$$

Moreover, there is equality in (1) if and only if x and y are linearly dependent, i.e. $x = \alpha y$ or $y = \alpha x$ for some $\alpha \in \mathbb{R}$.

Proof: Let $x, y \in V$

- If $x = 0$ or $y = 0$ the result is obvious.
- From now we assume that $\begin{cases} x \neq 0 \\ y \neq 0 \end{cases}$

Proof

We define
$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto \|y - tx\|^2 \end{aligned}$$

for $t \in \mathbb{R}$, $f(t) = \|y\|^2 - 2t\langle x, y\rangle + t^2\|x\|^2$

- Remark #1: $f(t)$ is a degree-2 polynomial (in t)
- Remark #2: $f(t) \geq 0$ for all t.

Hence the discriminant Δ of $f(t)$ is ≤ 0

$$\boxed{\Delta = 4\langle x, y\rangle^2 - 4\|x\|^2\|y\|^2} \leq 0 \rightarrow \text{we get Cauchy Schwarz !}$$

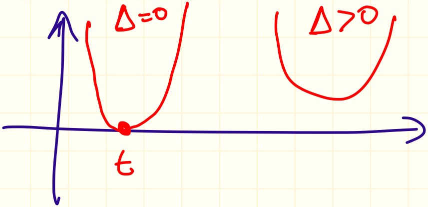
Proof

There is equality $|\langle x, y\rangle| = \|x\| \|y\|$ if and only if $\Delta = 0$

$\longrightarrow$ There exists some t such that

$$f(t) = 0$$
$$\|$$
$$\|y - tx\|^2 = 0$$ this means that $y - tx = 0$

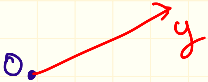
$$y = tx.$$



Questions?

Questions?

$$\cos\theta = \frac{x \cdot y}{\|x\|_2 \|y\|_2}$$



Orthogonal matrices

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ is called an *orthogonal matrix* if its columns are an orthonormal family.

Ex :

$$\mathrm{Id}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

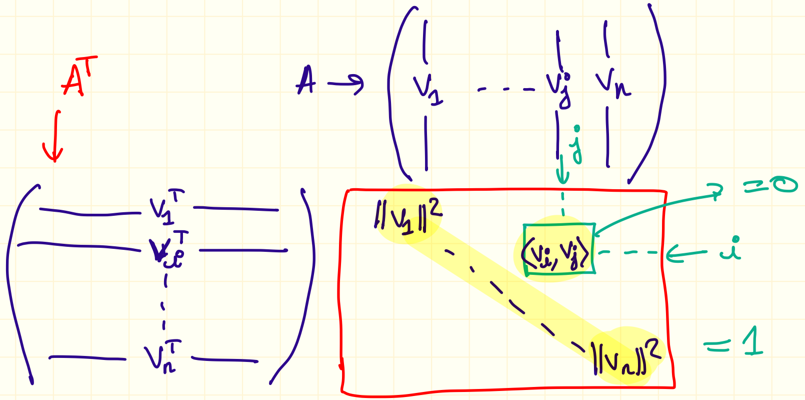
orthogonal matrices.

A proposition

Proposition

Let $A \in \mathbb{R}^{n\times n}$. The following points are equivalent:

1. A is orthogonal.
2. $A^\top A = \mathrm{Id}_n$. $\iff$ A invertible and $A^\top = A^{-1}$
3. $AA^\top = \mathrm{Id}_n$ $\iff$ ____________________



Orthogonal matrices & norm

Proposition

Let $A \in \mathbb{R}^{n \times n}$ be an orthogonal matrix. Then A preserves the dot product in the sense that for all $x, y \in \mathbb{R}^n$,

$$\langle Ax, Ay \rangle = \langle x, y \rangle.$$

In particular if we take $x = y$ we see that A preserves the Euclidean norm: $\|Ax\| = \|x\|$.