SDS 387
Linear Models

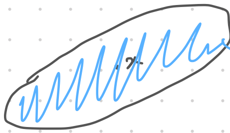
Fall 2025

Lecture 15 - Tue, Oct 21, 2025

Instructor: Prof. Ale Rinaldo

- Final project: send me by email a 1-page proposal by the end of next week. I will send a reminder.

- Projection onto a linear subspace wrt to $\langle x, y \rangle_{\Sigma} = x^T \Sigma y$ (different inner product) for some p.d. $\Sigma > 0$. This induces a different norm in $\mathbb{R}^d$ than the Euclidean one, which implies that distances are also different. A unit ball wrt to the induced norm $\|x\|_{\Sigma} = \sqrt{x^T \Sigma x}$ is a ellipsoid:



$\{x : \|x\|_{\Sigma} \leq r\}$

- Orthogonality is wrt $\langle \cdot, \cdot \rangle_{\Sigma}$. If $\mathcal{N}$ is a linear subspace of $\mathbb{R}^d$ spanned by columns of $\underset{d \times r}{U}$ $\left(\text{rank}(U) = r = \dim(\mathcal{N}) \right)$ then the oblique projection of

$x \in \mathbb{R}^d$ onto $\mathcal{N}$ wrt $\langle \cdot, \cdot \rangle_\Sigma$ is

$$P_\Sigma = U (U^T \Sigma U)^{-1} U^T \Sigma$$

Notice:

$\hookrightarrow$ · idempotent
· not symmetric

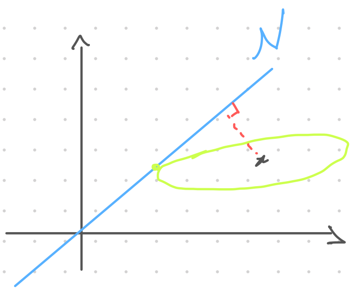
$$P_\Sigma U = U$$

and if $\langle x, u_i \rangle_\Sigma = 0 \quad \forall i = 1, \dots, k$ then

u_i: ith column of U

$$P_\Sigma x = 0$$

- Still we have a direct sum decomposition of any $x \in \mathbb{R}^d$:



$$x = x_\mathcal{N} + x_{\mathcal{N}^\perp} \qquad \text{where} \quad x_\mathcal{N} = P_\Sigma x$$

$$\langle x_\mathcal{N}, x_{\mathcal{N}^\perp} \rangle_\Sigma = 0 \qquad x_{\mathcal{N}^\perp} = (I - P_\Sigma) x$$

with $\mathcal{N}^\perp = \{ y \in \mathbb{R}^d : \langle x, y \rangle_\Sigma = 0 \ \ \forall x \in \mathcal{N} \}$

VECTOR / MATRIX NORMS

Recall that a norm over a vector space $\mathcal{X}$ is a function $\| \cdot \| : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$ s.t.

i) $\|a x\| = |a| \, \|x\| \quad \forall x \in \mathcal{X} \quad \forall a \in \mathbb{R}$

ii) $\|x\| = 0 \quad$ iff $\quad x = 0$

iii) $\|x+y\| \leq \|x\| + \|y\| \quad \forall x, y \in \mathcal{X}$

- A norm induces a notion of distance $d(x,y) = \|x-y\|$

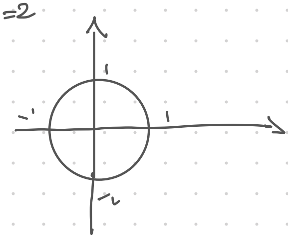
- In $\mathbb{R}^d$ for $p \geq 1$, the p-norm of a vector $x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$ is

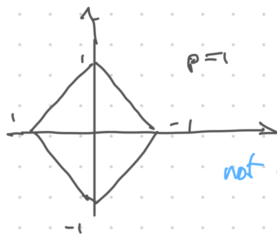
$$\|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{\frac{1}{p}}$$

For $p=2$ we get the Euclidean norm

Unit ball of $\|\cdot\|_p$

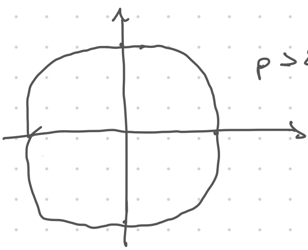
$p=2$



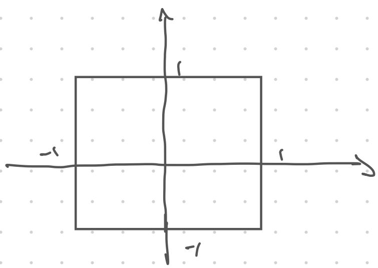


$p=1$

not drawn at scale



$p > 2$ and large



$\|x\|_\infty = \max_i |x_i|$

- In $\mathbb{R}^d$ all the $\|\cdot\|_p$ norms are equivalent

$1 \leq p \leq q \qquad \|x\|_q \leq \|x\|_p \leq d^{\frac{1}{p}-\frac{1}{q}} \|x\|_q$

$$\|x\|_1 \le \sqrt{d}\,\|x\|_2 \qquad \|x_1\| \le d\,\|x_\infty\|$$

$$\|x\|_2 \le \sqrt{d}\,\|x\|_\infty$$

Holder's ineq.

conjugate indexes

- $|\langle x, y\rangle| \le \|x\|_p \|y\|_q \qquad \frac{1}{p} + \frac{1}{q} = 1$

$$\le \|x\| \, \|y\| \qquad p = q = 2$$

- Remark: For the Euclidean norm:

$$\|x\|_2 = \|Ux\|_2 \qquad \text{any orthogonal matrix } U$$

- Matrix norms: $\underset{m\times n}{A}$ A matrix norm $|||\cdot|||$ is a norm on the space of $m \times n$ matrices if it satisfies all the properties of a norm and, in addition, is sub-multiplicative:

$$|||AB||| \le |||A|||\;|||B||| \qquad A, B \text{ conformal}$$

- One simple approach to define a matrix norm is to treat $\underset{m\times n}{A}$ as a vector and use a vector norm

For example, using the Euclidean norm:

Frobinius norm ←

$$\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2} = \sqrt{tr(AA^T)} = \sqrt{tr(A^TA)}$$

↓

Euclidean norm of its singular values

$$= \sqrt{\sum_i \sigma_i^2(A)}$$

↳ singular values of A

- $\|A\|_F$ is unitarily invariant.

$$\|A\|_F = \|UAV\|_F \qquad \underset{m\times m}{U} \quad \underset{n\times n}{V} \quad \text{orthogonal}$$

- p-Schatten norms $(p \geq 1)$:

$$\|A\|_p = \left(\sum_i \sigma_i^p(A) \right)^{1/p}$$

when $p=1$ this is called the nuclear norm of A.

- Hölder's ineq. for Schatten norms:

Inner product are matrices

$$\langle A, B \rangle = \operatorname{tr}(A^T B) \quad \text{this is an inner product}$$

$$\langle A, A \rangle = \|A\|_F^2$$

$$|\langle A, B \rangle| \leq \|A\|_p \|B\|_q \qquad \frac{1}{p} + \frac{1}{q} = 1$$

- $\|A\|_\infty = \max_{i,j} |A_{ij}|$ is another norm

- Operator norms

$\underset{m\times n}{A}$

let $\|\cdot\|_a$ and $\|\cdot\|_b$ two vector norm in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Then the

$$\|A\|_{a\to b} = \|A\|_{a,b} = \max_{\|x\|_a = 1} \|Ax\|_b$$

Most importan case $a = b = 2$ Then

$$\|A\|_{2,2} = \|A\|_{op} = \max_{\|x\|_2 = 1} \|Ax\|_2 = \sigma_{max}(A)$$

Los - Schatten norm

largest s.v. of A

- Property of operator norms:

$$\|Ax\|_2 \leq \|A\|_{op} \|x\|_2$$

In general

$$\|Ax\|_b \leq \|A\|_{a,b} \|x\|_a$$

Projection of a Random Variable

Van der Vaart chapter 11

- Let T and $\{S, s \in \mathcal{S}\}$ be a collection of random variables with finite second moments.

- A r.v. $\hat{S}$ is the L_2 projection of T onto $\mathcal{S}$ if $\hat{S}$ minimizes

not necessarily unique

$$s \in \mathcal{S} \longmapsto \mathbb{E}\left[(T-s)^2\right]$$

closed wrt to addition + scalar multiplication

- If $\mathcal{S}$ is a vector space then $\hat{S}$ is the projection of T onto $\mathcal{S}$ iff $T-\hat{S}$ is orthogonal to all r.v. $s \in \mathcal{S}$:

$$\mathbb{E}\left[(T-\hat{S}) \cdot s\right] = 0 \quad \forall s \in \mathcal{S}.$$

Aside: Think of the space of all r.v.'s with finite second moment as a Hilbert space wrt inner product

L_2

$$s_1, s_2 \in L_2 \rightarrow \langle s_1, s_2 \rangle = \mathbb{E}[s_1 \cdot s_2]$$

$$\langle T-\hat{S}, s \rangle = 0 \quad \forall s \in \mathcal{S}$$

Thm 11.1 (of Van der Vaart) $\hat{S}$ is the projection of T onto $\mathcal{S}$ iff i) $\hat{S} \in \mathcal{S}$ and ii) $\mathbb{E}\left[(T-\hat{S})s\right] = 0$

$\langle T-\hat{S}, s \rangle = 0 \quad \forall s \in \mathcal{S}$

This projection is unique (in the sense if $\hat{S}'$ is another projection then $\mathbb{P}(\hat{S} \neq \hat{S}') = 0$)

If $\mathcal{S}$ contains the constant functions then

$$\mathbb{E}[\hat{S}] = \mathbb{E}[T] \quad \text{and} \quad \mathrm{cov}(T - \hat{S}, S) = 0 \quad \forall S \in \mathcal{S}$$

Pf/

Assume orthogonality, i.e. $\mathbb{E}[(T - \hat{S})S] = 0 \quad \forall S \in \mathcal{S}$

Then for any $S \in \mathcal{S}$

$$\mathbb{E}[(T-S)^2] = \mathbb{E}[(T - \hat{S}) + (\hat{S} - S)]^2 =$$

$$= \mathbb{E}[(T-\hat{S})^2] + \underbrace{\mathbb{E}[(\hat{S}-S)^2]}_{\geq 0} + 2\underbrace{\mathbb{E}[(T-\hat{S})(\overbrace{S-\hat{S}}^{\in \mathcal{S}})]}_{= 0 \text{ by orthogonality}}$$

$$\geq \mathbb{E}[(T-\hat{S})^2]$$

We have an equality iff $\mathbb{E}[(S - \hat{S})^2] = 0$ iff $\mathbb{P}(S = \hat{S}) = 1$

⋮

Finish next time

⋮