SDS 387
Linear Models

Fall 2024

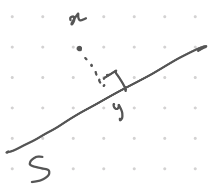
Lecture 13 - Tue, Oct 8, 2024

Instructor: Prof. Ale Rinaldo

- In $\mathbb{R}^d$, using standard inner product $\langle x, y \rangle = x^T y$, the orthogonal projection of a point $x \in \mathbb{R}^d$ <u>onto</u> a linear subspace S is the <u>unique</u> point $y \in S$ s.t.

$$y = \underset{z \in S}{\arg\min} \, \| x - z \|$$

orthogonality means that $y - x \in S^{\perp}$ (i.e. $\langle y - x, z \rangle = 0 \ \forall z \in S$)



Uniqueness and orthogonality follow from the fact that

$$\underbrace{x =}_{\text{unique decomposition}} x_S \underbrace{+}_{\text{direct sum}} x_{S^{\perp}} \qquad x_S \in S, \quad x_{S^{\perp}} \in S^{\perp}$$

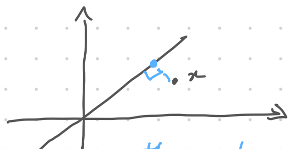
so if $\underset{x_S}{\overset{\shortparallel}{y}} \in S$ then $x - y = x_{S^{\perp}} \in S^{\perp}$ and

for any $y' \in S$

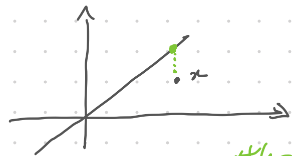
$$\|x-y\|^2 \le \|x-y\|^2 + \underbrace{\|y-y'\|^2}_{\ge 0} \underset{\text{by orthogonality}}{=} \|\overbrace{x-y}^{\in S^\perp} + \underbrace{y-y'}_{\in S}\|^2$$

$$= \|x-y'\|^2$$

- A projection does not have to be orthogonal



orthogonal



non orthogonal projection

in general a projection onto a linear subspace S is a mapping $T: \mathbb{R}^d \xrightarrow[\text{onto}]{} S$ s.t.

$$T \circ T(x) = T\big(T(x)\big) = T(x) \quad \left(\begin{array}{l}\text{it is an identity when} \\ \text{restricted to } S\end{array}\right)$$

$\rightarrow$ so $T(x) = x$ iff $x \in S$

A non-orthogonal projection is an oblique projection.

- Orthogonal projections are linear mappings!

For a linear subspace S in $\mathbb{R}^d$ of dimension $1 \le r \le n$ the orthogonal projection of x onto S is given by

$$\underset{d \times d}{P} x = y \in S$$

where P is a projector or projection matrix that satisfies these defining properties

definition of a projection ← i) $P^2 = P$ (idempotent)

orthogonality ← ii) P is symmetric

In fact, any $d \times d$ in $\mathbb{R}^d$ with these properties is a projector

- P is positive semi-definite   Exercise
- Projectors are unique (i.e. $Px = Qx$ $\forall x$ implies $P = Q$)
  Exercise
- Explicit expression for P: let $\underset{d \times r}{A}$ s.t. $C(A) = S$
  ↓ columns form a basis for S

Then

$$P = A \underbrace{(A^T A)^{-1}}_{\text{invertible}} A^T$$

if columns of A are orthonormal $A = [a_1 \ldots a_r]$, then

$$P = AA^T \quad \text{and} \quad Px = \sum_{i=1}^{r} a_i \langle a_i, x \rangle$$

↓ linear combinations of the a_i's with coeff. $\langle a_i, x \rangle$

Also $\quad \|Px\|^2 = \sum_{i=1}^{r} (\langle a_i, x \rangle)^2$

- P has r non-zero eigenvalues, all equal to 1.

- Projections wrt to a different inner product:

Σ pd $\quad \langle x, y \rangle_{\Sigma} = x^T \Sigma y$

This induce the norm $\|x\|_{\Sigma} = \sqrt{x^T \Sigma x}$ and distance $\|x-y\|_{\Sigma}$

level set of $\|x_0\|$ $\quad \rightarrow$ level set of $\|x_0\|_{\Sigma}$

Now orthogonality is wrt to $\langle \cdot, \cdot \rangle_{\Sigma}$

If S is a linear subspace of $\mathbb{R}^d$ spanned by orthonormal columns of $\underset{d \times r}{U}$ the the orthogonal projector wrt to $\langle \cdot, \cdot \rangle_{\Sigma}$ is

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

$\rightarrow$ not symmetric

We can see this because

$P_{\Sigma} U = U \quad$ or $\quad P_{\Sigma} u_i = u_i \rightarrow$ ith column of U

and if $\langle x, u_i \rangle_{\Sigma} = 0$ all i then

$$P_\Sigma x = 0$$

Of course orthogonality wrt to $\langle \cdot, \cdot \rangle$ does not hold.

Also any $x \in \mathbb{R}^d$ can be written as

$$x = x_S + x_{S^\perp} \qquad x_S = P_\Sigma x$$

$$x_{S^\perp} = (I - P_\Sigma) x$$

$$S^\perp = \{ y : \langle y, x \rangle_\Sigma = 0 \ \forall x \in S \} \qquad \langle x_S, x_{S^\perp} \rangle_\Sigma = 0$$

VECTOR / MATRIX NORMS

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

i) $\| \alpha x \| = |\alpha| \, \| x \|$

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

iii) $\| x + y \| \leq \| x \| + \| y \|$

For a point $x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{R}^d$ and $p \geq 1$

its p-norm is

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

and its ∞-norm is $\| x \|_\infty = \lim_{p \to \infty} \| x \|_p = \max_i |x_i|$

In $\mathbb{R}^d$ these are all equivalent norms:

if $1 \leq p \leq q$ then

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

$$\|x\|_1 \leq \sqrt{d}\, \|x\|_2 \qquad \|x_1\| \leq d\, \|x\|_\infty$$

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

Holder inequality if $x, y \in \mathbb{R}^d$

$$|\langle x, y\rangle| \leq \sum_{i=1}^{d} |x_i y_i| \leq \|x\|_p \|y\|_q \qquad \frac{1}{p} + \frac{1}{q} = 1$$

- the L_2 norm is the one induced by standard inner product $\|x\| = \sqrt{x^T x}$. This implies that it is unitarily invariant:

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

- MATRIX NORMS

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

$$|||\tilde{A}B||| \leq |||\tilde{A}|||\,|||B|||$$

- Simple approch: Treat $A_{m \times n}$ as a vector in $\mathbb{R}^{mn}$ and apply any vector norm. Example

$$\|A\|_\infty = \max_{i,j} |A_{i,j}|$$

Norm induced by inner product $\langle A, B \rangle = tr(A^T B)$

"Natural norm"

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

Frobenius norm

$$= \sqrt{\sum_{i=1}^{r} \sigma_i^2 (A)}$$

r → rank (A)

↳ singular values of A

Also $\|\cdot\|_F$ is unitarily invariant:

$$\|A\|_F = \|U A V\|_F \qquad U_{m \times m},\ V_{n \times n} \text{ orthogonal}$$

- Another type of matrix norm is the p-Schatten norm

the p-schatten norm of $A_{m \times n}$ is $\quad p \geq 1$

$$\|A\|_p = \left(\sum_{i=1}^{\text{rank}(A)} |\sigma_i(A)|^p \right)^{1/p}$$

↳ singular values of A

→ $\|\cdot\|_F$ is the 2-Schatten norm

the 1-Schatten norm is called the nuclear norm of A

→ the ∞ - schatten is called the operator norm of A

Holder's inequality holds:

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

- Finally, we can define a third type of norm for matrices using the notion of operator norm.

If $1 \leq p, q \leq \infty$ the p,q operator norm of $A_{m \times n}$ is

$$\|A\|_{p,q} = \max_{\substack{x \in \mathbb{R}^n \\ \|x\|_p = 1}} \|Ax\|_q$$

idea comes from this fact: if $x \in \mathbb{R}^d$

$$\|x\|_p = \max_{\substack{y \in \mathbb{R}^d \\ \|y\|_q = 1}} x^T y$$

$\frac{1}{p}+\frac{1}{q}=1$

Duality btw L_p and L_q norm

Examples

This gives us the spectral or operator norm of A:

$$\|A\|_{op} = \max_{\|x\|=1} \|Ax\| = \sqrt{\lambda_{max}(A^T A)}$$

$p=q=2$

$$= \sigma_{max}(A)$$

largest singular value

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

Exercise!