SDS 387
Linear Models

Fall 2025

Lecture 13 - Tue, Oct 14, 2025

Instructor: Prof. Ale Rinaldo

Some useful linear algebra references:

- Matrix Analysis by Horn & Johson
- Matrix Analysis by Bhatia

} Very rigorous & comprehensive

- Matrix Computations by Golub } Algorithmic focus
- Matrix Perturbation Theory by Sun & Stewart
- Linear Algebra Done Right by Axler
- Introduction to Applied Linear Algebra by Boyd

} Introductory references available online

- Appendix to Plane Answers to ~~Simple~~ Complex Questions by Christensen (available online from Springerlink)

- For the statistics/ML results about linear models, we will use next the book: Learning Theory from First Principles by Francis Bach (available online)

- Matrix Algebra: Theory, Computations & Applications in statistics by J.E. Gentle (Springerlink)

- We will be working in $\mathbb{R}^d$
- Vector or linear space (over $\mathbb{R}$): a collection of points in $\mathbb{R}^d$ closed wrt scalar multiplication and addition (and has a 0 (zero) element).

  M vector space in $\mathbb{R}^d$: $x, y \in M \Rightarrow$
  $ax + by \in M$ $\forall a, b \in \mathbb{R}$

- A linear subspace N of a linear space M is a subset of M that is also a linear space. In $\mathbb{R}^d$

$$\left\{ x = \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_d \end{bmatrix} : \begin{array}{c} x_{k+1} = x_{k+2} = \cdots = x_d = 0 \\ 1 < k < d \end{array} \right\}$$

  is a linear subspace of $\mathbb{R}^d$

- Geometrically linear subspaces in $\mathbb{R}^2$ are lines through the origins

- A finite subset of M $\{v_1, \dots, v_r\}$ is a set of linearly independent vectors when

$$\underbrace{\sum_{i=1}^{n} \underset{\in \mathbb{R}}{a_i} v_i}_{\text{linear combination}} = 0 \rightarrow a_1 = a_2 = \cdots = a_r = 0$$

- A set of linearly ind. vectors $\{v_1 , \dots , v_r\}$ span a subspace N (of $\mathbb{R}^d$) when every $x \in N$ can be written as a linear combinations of the v_i's.

In this case the v_i's are called a basis of N. Bases are not unique but the number of elements in each basis is unique and is called the <u>dimension</u> of N

<u>Fact</u>: if $\{v_1, \dots, v_r\}$ form a basis for N then $\forall x \in N$ $\underbrace{\exists!}_{\text{there exist a \underline{unique} set}} a_1, \dots, a_r$ s.t.

$$x = \sum_{i=1}^{r} a_i v_i$$

- if N_1 and N_2 are subspaces, so are

$$N_1 + N_2 = \left\{ x \in \mathbb{R}^d : \begin{array}{l} x = x_1 + x_2 \\ \text{some } x_1 \in N_1 \text{ and } x_2 = N_2 \end{array} \right\}$$

$$N_1 \cap N_2$$

- What about $N_1 \cup N_2$? no

$\mapsto$ zero

- if $N_1 \cap N_2 = \{0\}$ then

dimension / rank $(N_1 + N_2) = \dim(N_1) + \dim(N_2)$

- In $\mathbb{R}^d$, there is a notion of inner product

a function $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d$ that

is symmetric $(\langle x, y \rangle = \langle y, x \rangle)$,

linear $(\langle a(x+y), bz \rangle = ab \langle x, z \rangle + ab \langle y, z \rangle)$

is positive definite:

$\downarrow$

$\langle x, x \rangle > 0$

In $\mathbb{R}^d$ for $x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix}$

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{d} x_i y_i$$

$\downarrow$

This gives the Euclidean norm:

$$\|x\| = \sqrt{\langle x, x \rangle}$$

$\hookrightarrow$ magnitude of x

- There can be other choices of inner products: if $\underset{d \times d}{A} \succ 0$ positive definite

$$\langle x, y \rangle_A := x^T A y$$

is an inner product

$$x^T A x = \sum_{i,j} A_{i,j} x_i x_j > 0$$

$$\left(\begin{array}{c} \sum_{i,j} A_{i,j}\, x_i x_j > 0 \\ \text{for all} \\ x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \end{array} \right)$$

- Inner products allow to define the notion of <u>orthogonality</u>: x and y are orthogonal when $x^T y = \langle x, y \rangle = 0$.

- An orthogonal basis is a basis consisting of orthogonal vectors. It is orthonormal when the basis elements have unit norm i.e. $\|v_i\| = 1$ all basis elements $v_1, \ldots, v_r$

- If $v_1, \ldots, v_r$ is a basis for a r-dimensional linear subspace you can always find an orthonormal basis. This process is called the Gram-Schmidt orthogonalization:

$$y_1 = \frac{v_1}{\|v_1\|}$$

For $i = 2, \ldots, r$ let $\begin{cases} w_i = v_i - \sum_{j=1}^{i-1} \langle v_i, w_j \rangle w_j \\ y_i = \frac{w_i}{\|w_i\|} \end{cases}$

Then $y_1, \ldots, y_r$ is an orthonormal basis of $\mathcal{N}$.

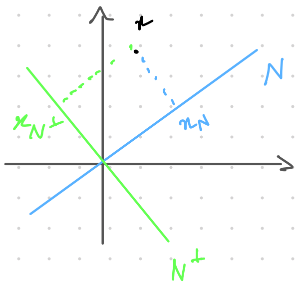
- If N is a linear subspace of $\mathbb{R}^d$, its orthogonal complement (in $\mathbb{R}^d$) is the linear subspace

$$N^\perp = \{x \in \mathbb{R}^d : \langle x, y \rangle = 0, \ \forall y \in N\}$$

- <u>Fact</u>: $N \cap N^\perp = \{0\}$

- Any vector $x \in \mathbb{R}^d$ can be written uniquely as

$$\underbrace{x = x_N + x_{N^\perp}}_{\text{direct sum}} \qquad \text{where } x_N \in N, \ x_{N^\perp} \in N^\perp$$



by definition $\langle x_N, x_{N^\perp} \rangle = 0$

$\downarrow$

$$\mathbb{R}^d = N + N^\perp \quad \text{and}$$

$$d = \dim(\mathbb{R}^d) = \dim(N) + \dim(N^\perp)$$

- <u>Fact</u>: $(N_1 \cap N_2)^\perp = N_1^\perp + N_2^\perp$

MATRICES

- In $\mathbb{R}^d$ a vector is a 1-dim array of numbers.

A matrix is a 2-dim array:

$$A = (A_{ij})_{\substack{i=1\ldots m \\ j=1,\ldots,n}} \in \mathbb{R}^m \times \mathbb{R}^n$$

size $\leftarrow m \times n$

- Set of matrices are closed under scalar multiplication and addition (provided that they are of the same size)

- Notion of <u>product</u>: $\underbrace{\underset{m\times n}{A}\ \underset{n\times k}{B}}_{\text{conformal}} = \underset{m\times k}{C}$

$$C_{i,j} = \sum_{\ell=1}^{n} A_{i,\ell} B_{\ell,j}$$

- Big issue: non-commutative. In general

$$AB \neq BA$$

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

$C(A) = R(A)$: linear subspace of $\mathbb{R}^m$ spanned by columns of A
column range of A

kernel (A)
nullspace (A) : linear subspace in $\mathbb{R}^n$ $\{x \in \mathbb{R}^n : Ax = 0\}$

- Transpose of $\underset{m\times n}{A} = (A_{i,j})$ is the matrix

$$\underset{n\times m}{A^T} = (A_{j,i})$$

- $(AB)^T = B^T A^T$

- $\underset{n\times n}{A}$ is symmetric when $A = A^T$

- $\underset{n\times n}{A}$ square matrix is diagonal when $A_{i,j} = 0$
$\forall i \neq j$

- I_n diagonal matrix with unit elements along diagonal

$$I_m \underset{m\times n}{A} = A I_n = A$$

- The inverse of $\underset{n\times n}{A}$ is the matrix A^{-1} s.t.

does not always exist!

$$A^{-1}A = AA^{-1} = I_n$$

The inverse is unique!

- Note : $(AB)^{-1} = B^{-1}A^{-1}$

- if $\underset{n\times n}{A}$ is invertible it is said to be non-singular

This happens iff $\text{rank}(A) = n$

# of linearly indep. rows or columns

- Fact :

if $\dim\left(R(\underset{n\times n}{A})\right) = r$ then

$$\dim(\text{null}(A)) = n - r$$

- A matrix $\underset{n\times n}{U}$ is orthogonal when $\xrightarrow{}$ has orthonormal columns

$$UU^T = I_n = UU^T$$

- The trace of $\underset{n\times n}{A}$ is $tr(A) = \sum_{i=1}^{n} A_{ii}$

$tr(\cdot)$ is a linear function
$tr(aA + B) = a\,tr(A) + tr(B)$

$tr(\cdot)$ has a nice property

$$tr(ABC) = tr(CAB) = tr(BCA)$$

$$\neq tr(ACB)$$

- In fact one can use the $tr(\cdot)$ to define an inner product over space of squared matrices:

$$\langle B, A\rangle = \langle A, B\rangle = tr(AB^T) = tr(B^TA)$$

SPECTRAL PROPERTIES

$\underset{n\times n}{A}$: $\lambda \in \mathbb{C}$ is an eigenvalue of A
if $(A - \lambda I_n)$ is singular

- We can find all the eigenvalues of A by solving a polynomial equation:

$$\det(A - \lambda I) = 0$$

- If $\lambda \in \mathbb{R}$ is an eigenvalue of A its corresponding eigenvector is $x \in \mathbb{R}^n$, $x \neq 0$ s.t.

$$Ax = \lambda x \quad (\text{i.e. } (A - \lambda I)x = 0)$$

- If A has $r \leq n$ eigenvalues $\lambda_1, \ldots, \lambda_r$ then

$$\lambda \mapsto \det(A - \lambda I) = \prod_{j=1}^{r} (\lambda_j - \lambda)^{\mu(\lambda_j)}$$

$\mu(\lambda_j)$: algebraic multiplicity of λ_j

- If λ is an eigenvalue of A, the dimension of the linear subspace $\text{null}(A - \lambda I_n)$ is the geometric multiplicity of λ

- Simple eigenvalue: have multiplicity 1

- $\underset{n \times n}{A}$ is symmetric then eigenvalues are real and there exists a variational characterization of the eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

$$\lambda_i = \max_{\substack{N \subseteq \mathbb{R}^n \\ \dim(N) = i}} \ \min_{\substack{x \in N \\ \|x\| = 1}} x^T A x$$

Courant - Fisher characteristics

$$= \min_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T) = n-i+1}} \ \max_{\substack{x \in T \\ \|x\| = 1}} x^T A x$$

$\Downarrow$

$$\lambda_1 = \lambda_{max}(A) = \max_{\substack{x \in \mathbb{R}^n \\ \|x\| = 1}} x^T A x$$

$$\lambda_n = \lambda_{min}(A) = \min_{\substack{x \in \mathbb{R}^n \\ \|x\| = 1}} x^T A x$$

$$\lambda_{n-i} = \min_{\|x\| = 1} x^T A x$$

$$x \perp S_{n-i+1} = \operatorname{span}(v_n, \ldots, v_{n-i})$$

$\perp$ ↓ orthogonal

$(v_n, \ldots, v_{n-i})$ ↓ eigenvectors

- If $\operatorname{rank}(A) = r < n$ then $\lambda_{r+1}, \ldots, \lambda_n = 0$

- Spectral Theorem $\underset{n\times n}{A}$ $\text{rank}(A) = r$

Then

$$A = \underset{n\times r}{U} \underset{r\times r}{\Lambda} U^T$$

U has orthonormal columns spanning $C(A)$

$$= \sum_{i=1}^{r} \lambda_i \underbrace{u_i u_i^T}_{n\times n \text{ of rank } 1}$$

u_i: ith column of U

λ_i: ith eigenvalues

Λ = diagonal matrix containing eigenvalues

↓

λ_i, u_i pair of eigenvalue eigenvector

Eigenspace of eigenvalue λ:

$\text{null}((A - \lambda I))$

symmetric $A \succeq 0$

- $\underset{n\times n}{A}$ is positive (semi) definite when

$$\underbrace{x^T A x}_{\sum_{i,j} A_{i,j} x_i x_j} \geq 0 \quad \forall x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

negative (semi) definite if $x^T A x \leq 0$ $\forall x \in \mathbb{R}^n$

$A \preceq 0$

- If Σ is the covariance matrix of a random vector $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ (i.e. $\Sigma_{i,j} = \text{cov}[X_i, X_j]$)

then $\Sigma_1 \succeq 0$, because

$$c^T \Sigma_1 c = \mathrm{Var}[c^T X] \geq 0, \forall c \in \mathbb{R}^d$$