SDS 387
Linear Models

Fall 2024

Lecture 11 - Tue, Oct 1, 2024

Instructor: Prof. Ale Rinaldo

Some useful 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 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)

LINEAR ALGEBRA RECAP

- We will be working in $\mathbb{R}^d$ but much of what we say holds in more general spaces

- A **vector space** (over $\mathbb{R}$) or linear: a set closed wrt scalar multiplication and addition. M vector space

  It has a zero element: $x \in M \to \alpha x \in M$, $\alpha \in \mathbb{R}$

  $x + 0 = x$ → zero element

  $x, y \in M \to x+y \in M$

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

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

  In $\mathbb{R}^2$ a linear subspaces are lines through the origins

- A (finite) subset of M $\{v_1, \dots, v_d\}$ is a set of linearly independent vectors or points if

$$\sum_{i=1}^{d} \alpha_i v_i = 0 \longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_d = 0$$

  ↓ linear combination

- A set of linearly independent vectors $\{v_1, \dots, v_k\}$ spans a subspace N of M when every $x \in N$ can be written as a linear combination of the v_i's.

- In this case, $\{v_1, \dots, v_k\}$ is called a <u>basis</u> of N. Bases are not unique, but the number of elements in each basis is the same and it is called the dimension or rank of the subspace.

- <u>Facts</u>: if $\{v_1, \dots, v_k\}$ is a basis for N then $\forall x \in N \quad \exists ! \; \alpha_1, \dots, \alpha_k \in \mathbb{R}$

  there exists — unique

  s.t. $\quad x = \sum_{i=1}^{k} \alpha_i v_i$

- If N_1 and N_2 are subspaces, so is

  $$N_1 + N_2 = \{x : x = x_1 + x_2 \quad x_1 \in N_1 \quad x_2 \in N_2\}$$

  and

  $$N_1 \cap N_2$$

- What about $N_1 \cup N_2$? No

- If $N_1 \cap N_2 = 0$ then

  $$\text{rank}(N_1 + N_2) = \text{rank}(N_1) + \text{rank}(N_2)$$

- In $\mathbb{R}^d$ (and in many other spaces) we have an inner product, a function on $\mathbb{R}^d \times \mathbb{R}^d$ that is symmetric $(\langle x, y \rangle = \langle y, x \rangle)$, linear

$\left(\langle \alpha x, \beta y \rangle = \alpha\beta \langle x, y \rangle \qquad \alpha, \beta \in \mathbb{R} \right)$

and positive definite $\langle x, x \rangle \geq 0$.

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

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

This gives the Euclidean norm $\|x\| = \sqrt{\langle x, x \rangle}$

- On the same space you can define more than one inner product. (Example: if $\Sigma \succ 0$

$$\langle x, y \rangle_\Sigma = x^T \Sigma y = \sum_{i,j} x_i y_j \Sigma_{i,j} \Big)$$

- Inner products allow to define orthogonality:

  x and y are orthogonal when $\langle x, y \rangle = 0$

- An orthogonal basis is a basis consisting of orthogonal vector. An the basis is orthonormal if it is orthogonal and all its elements have unit norm (the norm induced by inner product).

- If $v_1, \dots, v_k$ is a basis for some subspace there always exists an orthonormal basis that can be constructed using $v_1, \dots, v_k$. This process is known as Gram-Schmidt orthogonalization)

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

for $i = 2, \dots, k$ 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 $y_1, \dots, y_k$ are an orthonormal basis HW

- If S is a subspace of M, the orthogonal complement of S (in M) is the linear subspace
$$S^\perp = \{ x \in M : \langle x, y \rangle = 0 \quad \forall y \in S \}$$

- Fact: $S \cap S^\perp = \{0\}$

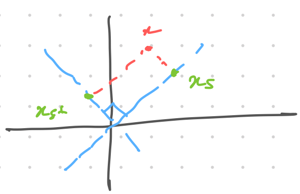
↳ Any vector $x \in M$ can be written uniquely as
$$x = x_S + x_{S^\perp} \quad \text{where } x_S \in S, \; x_{S^\perp} \in S^\perp$$

direct sum

of course $\langle x_S, x_{S^\perp} \rangle = 0$

As a result $M = S + S^\perp$ and

$$\operatorname{rank}(M) = \operatorname{rank}(S) + \operatorname{rank}(S^\perp)$$



Fact: If S_1 and S_2 are subspaces
$$(S_1 \cap S_2)^\perp = S_1^\perp + S_2^\perp$$

MATRICES

- In $\mathbb{R}^d$ a vector is a 1-dim array. A matrix is a 2-dim array:

$$\underset{m \times n}{A} = \left(A_{i,j}\right)_{\substack{i=1\dots m \\ j=1\dots n}} \begin{array}{l} \text{rows} \\ \text{columns} \end{array} \in \mathbb{R}^m \times \mathbb{R}^n$$

- Set of matrices is closed wrt scalar multiplication and addition (if the matrices have same size)

- Notion of product:

$$\underbrace{\underset{m \times n}{A} \quad \underset{n \times k}{B}}_{\text{conformal}} = \underset{m \times k}{C} \qquad C_{i,j} = \sum_{e=1}^{n} A_{i,e} B_{e,j}$$

Big issue: non commutativity. In general

$$AB \neq BA$$

- $\underset{m \times n}{A}$ $\quad R(A)$: linear subspace of $\mathbb{R}^m$ spanned by columns of A

Kernel(A)
nullity(A) : linear subspace of $\mathbb{R}^n$ of the form $\{x \in \mathbb{R}^n : Ax = 0\}$
nullspace(A)

$$\sum_{i=1}^{n} A_i x_i$$

A_i ith column of A

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

$\underset{n\times m}{A^T} = (A_{j,i})$. Note: $(AB)^T = B^T A^T$

A matrix $\underset{n\times n}{A}$ is <u>symmetric</u> when $A = A^T$

- $\underset{m\times n}{A}$ is square when $m=n$. A square matrix is

diagonal when all elements $A_{i,j} = 0 \quad \forall i \neq j$

↓ I_d = diagonal matrix with $I_{d\,i,i} = 1 \quad \forall i$

identity $\quad I_n \underset{m\times n}{A} = \underset{m\times n}{A} I_n = A$

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

$A^{-1}A = AA^{-1} = I_n$ $\qquad$ ↓ unique

<u>Note</u>: $(AB)^{-1} = B^{-1}A^{-1}$

- If $\underset{d\times d}{A}$ has an inverse it is said to be non-singular

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

↓ # of linearly independent columns or rows

- If $\text{rank}\big(\underbrace{C(A)}_{\text{range or column space}}\big) = r$ then $\text{rank}(\text{null}(\underset{n\times n}{A})) = n-r$

- A matrix $\underset{n\times n}{U}$ is orthogonal when its columns are orthonormal vectors. Then

$$U^T U = I_n = U U^T$$

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

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

  $tr(\cdot)$ has cyclic property

  $$tr(ABC) = tr(CAB) = tr(BCA) \neq tr(ACB)$$

- In fact you can define an inner product over square matrices

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

SPECTRAL PROPERTIES

Let $\underset{n\times n}{A}$. Number λ is an eigenvalue of A if $(A - \lambda I_n)$ is singular

A corresponding eigenvector is a vector $x \in \mathbb{R}^n$ s.t. $\underset{x \neq 0}{}$ $Ax = \lambda x$ $\quad ((A - \lambda I)x = 0)$

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

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

If A has $r \leq n$ eigenvalues $\lambda_1, \dots, \lambda_r$

then

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

algebraic multiplicity of λ_j → positive integer

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

algebraic multiplicity

- Simple eigenvalues are those with multiplicity 1.

- If $\underset{n\times n}{A}$ is symmetric then (the eigenvalues are real) there exists a variational characterization of its eigenvalues: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ possibly repeating

Then

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

linear subspace

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

linear subspace