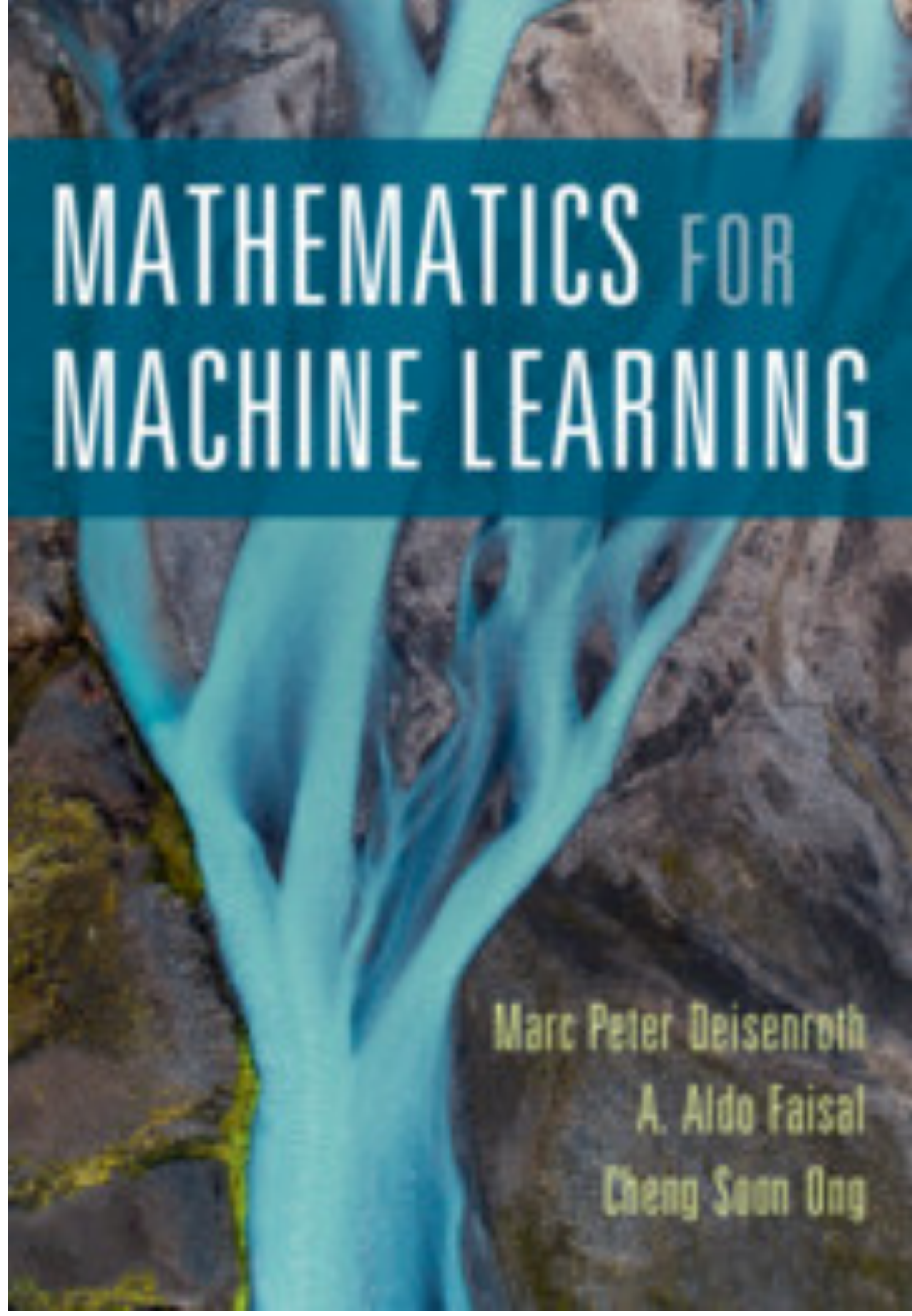
2. Recap: Linear Algebra

EECE454 Introduction to
Machine Learning Systems

2023 Fall, Jaeho Lee

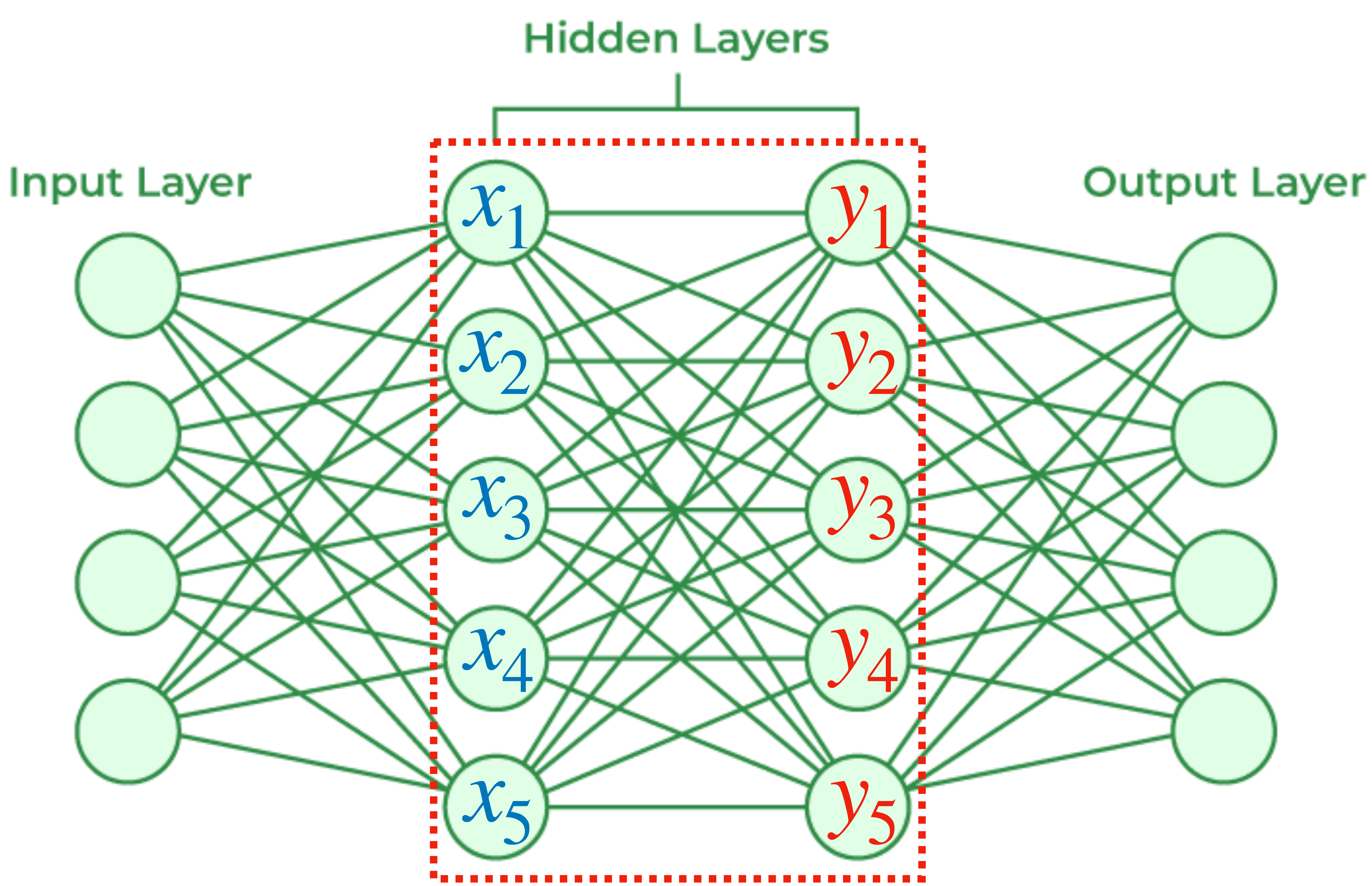
Disclaimer



- Use this slide as a guide for self-study!
- **Reference.**
  - MML book: Chapter 1 — Chapter 6
  - *Dive into Deep Learning*: Sec 2.3.—2.6.
  - Stanford CS229
    https://cs229.stanford.edu/lectures-spring2022/cs229-linear_algebra_review.pdf
  - 3Blue1Brown Youtube "Linear Algebra"
    https://www.3blue1brown.com/topics/linear-algebra

Why Linear Algebra?

- We use matrices to model the relationship between multi-dimensional input and multi-dimensional output.



model parameter (or "internal state")

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} \\ w_{21} & w_{22} & w_{23} & w_{24} & w_{25} \\ w_{31} & w_{32} & w_{33} & w_{34} & w_{35} \\ w_{41} & w_{42} & w_{43} & w_{44} & w_{45} \\ w_{51} & w_{52} & w_{53} & w_{54} & w_{55} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

Vectors and Matrices

Symbol	Typical meaning
$a, b, c, \alpha, \beta, \gamma$	Scalars are lowercase
$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$	Vectors are bold lowercase
$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$	Matrices are bold uppercase
$\boldsymbol{x}^\top, \boldsymbol{A}^\top$	Transpose of a vector or matrix
$\boldsymbol{A}^{-1}$	Inverse of a matrix
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle$	Inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$
$\boldsymbol{x}^\top \boldsymbol{y}$	Dot product of $\boldsymbol{x}$ and $\boldsymbol{y}$
$B = (\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$	(Ordered) tuple
$\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3]$	Matrix of column vectors stacked horizontally
$\mathcal{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$	Set of vectors (unordered)
$\mathbb{Z}, \mathbb{N}$	Integers and natural numbers, respectively
$\mathbb{R}, \mathbb{C}$	Real and complex numbers, respectively
$\mathbb{R}^n$	n-dimensional vector space of real numbers

Symbol	Typical meaning
$\forall x$	Universal quantifier: for all x
$\exists x$	Existential quantifier: there exists x
$a := b$	a is defined as b
$a =: b$	b is defined as a
$a \propto b$	a is proportional to b, i.e., $a = \text{constant} \cdot b$
$g \circ f$	Function composition: "g after f"
$\iff$	If and only if
$\implies$	Implies
$\mathcal{A}, \mathcal{C}$	Sets
$a \in \mathcal{A}$	a is an element of set $\mathcal{A}$
$\emptyset$	Empty set
$\mathcal{A} \backslash \mathcal{B}$	$\mathcal{A}$ without $\mathcal{B}$: the set of elements in $\mathcal{A}$ but not in $\mathcal{B}$

Symbol	Typical meaning
$\boldsymbol{I}_m$	Identity matrix of size $m \times m$
$\mathbf{0}_{m,n}$	Matrix of zeros of size $m \times n$
$\mathbf{1}_{m,n}$	Matrix of ones of size $m \times n$
$\boldsymbol{e}_i$	Standard/canonical vector (where i is the component that is 1)
$\dim$	Dimensionality of vector space
$\mathrm{rk}(\boldsymbol{A})$	Rank of matrix $\boldsymbol{A}$
$\mathrm{Im}(\Phi)$	Image of linear mapping Φ
$\ker(\Phi)$	Kernel (null space) of a linear mapping Φ
$\mathrm{span}[\boldsymbol{b}_1]$	Span (generating set) of $\boldsymbol{b}_1$
$\mathrm{tr}(\boldsymbol{A})$	Trace of $\boldsymbol{A}$
$\det(\boldsymbol{A})$	Determinant of $\boldsymbol{A}$
$\lvert \cdot \rvert$	Absolute value or determinant (depending on context)
$\lVert\cdot\rVert$	Norm; Euclidean, unless specified
$\boldsymbol{x} \perp \boldsymbol{y}$	Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal
V	Vector space
$V^{\perp}$	Orthogonal complement of vector space V

Quiz # 1

Let there be a vector $\mathbf{x} \in \mathbb{R}^n$ (we use boldcase, usually)
This is ...

(a)

$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n]$$

(b)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

Answer

Let there be a vector $\mathbf{x} \in \mathbb{R}^n$.
This is ...

(a)

$$\mathbf{x}^\top = [x_1 \quad x_2 \quad \cdots \quad x_n]$$

We call this "$\mathbf{x}$ transposed"

(b)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

Quiz # 2

Let there be a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. (bold uppercase)

This is ...

(a)

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{1n} \\ & & \cdots & \\ a_{m1} & a_{m2} & \cdots & x_{mn} \end{bmatrix}$$

(b)

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ & & \cdots & \\ a_{1n} & a_{2n} & \cdots & x_{mn} \end{bmatrix}$$

Answer

Let there be a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. (bold uppercase)

This is ...

m rows and *n* columns...

(a)

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{1n} \\ & & \cdots & \\ a_{m1} & a_{m2} & \cdots & x_{mn} \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} - & \mathbf{a}_1^\top & - \\ & \cdots & \\ - & \mathbf{a}_m^\top & - \end{bmatrix}$$

Multiplications

Vector products

- Two types: Inner / Outer.

Inner product (a.k.a. dot product)

$$\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

alternate notation
(only called inner, more general)

$$\langle \mathbf{x}, \mathbf{y} \rangle$$

Outer product

$$\mathbf{x}\mathbf{y}^\top = \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_n \\ & \cdots & \\ x_m y_1 & \cdots & x_m y_n \end{bmatrix}$$

Not very frequent though.

Matrix-Vector Multiplications

- Performing many inner products with row vectors.
    - or, we are summing many column vectors

$$\mathbf{W}\mathbf{x} = \begin{bmatrix} - & \mathbf{w}_1^\top & - \\ & \cdots & \\ - & \mathbf{w}_m^\top & - \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{w}_1^\top \mathbf{x} \\ \cdots \\ \mathbf{w}_m^\top \mathbf{x} \end{bmatrix}$$

Matrix-Vector Multiplications

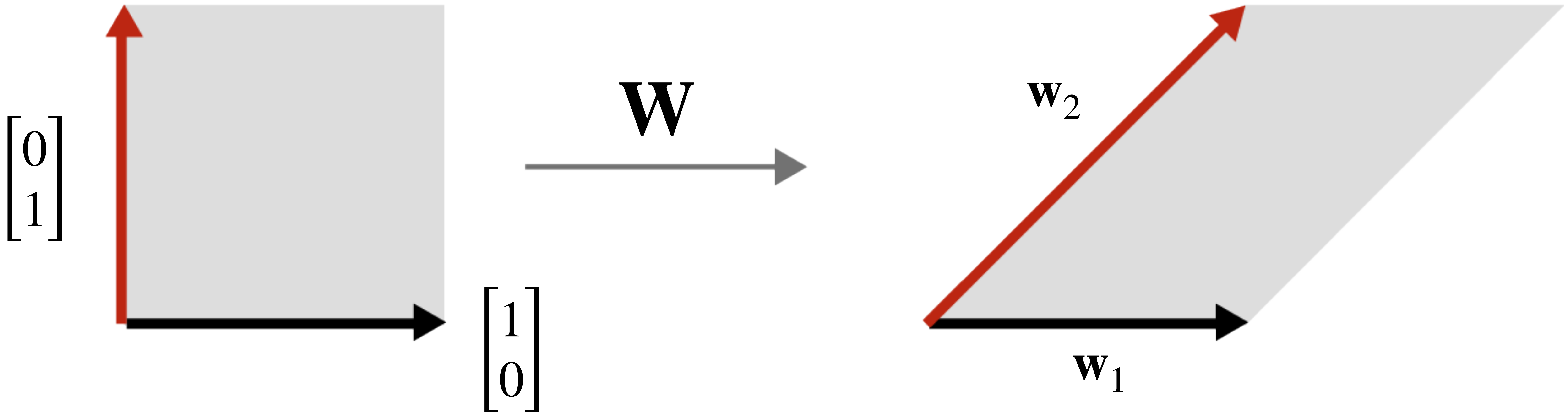
- Performing many inner products with row vectors.
  - or, a weighted sum of column vectors

$$\mathbf{W}\mathbf{x} = \begin{bmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ | & & | \end{bmatrix} \mathbf{x} \qquad = x_1\mathbf{w}_1 + \cdots + x_n\mathbf{w}_n$$

Physical Meaning ... System perspective

The matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be viewed as axis transformation

$$\mathbf{W}\begin{bmatrix}1\\0\\\cdots\\0\end{bmatrix} = \mathbf{w}_1 \qquad \cdots \qquad \mathbf{W}\begin{bmatrix}0\\0\\\cdots\\1\end{bmatrix} = \mathbf{w}_n$$



Matrix-Matrix Multiplications

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} = \mathbb{R}^{n \times p}$.
- Performing $m \times p$ inner products

$$\mathbf{AB} = \begin{bmatrix} - & \mathbf{a}_1^\top & - \\ & \cdots & \\ - & \mathbf{a}_m^\top & - \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{b}_1 & \cdots & \mathbf{b}_p \\ | & & | \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \mathbf{b}_1 & \cdots & \mathbf{a}_1^\top \mathbf{b}_p \\ \cdots & \cdots & \cdots \\ \mathbf{a}_m^\top \mathbf{b}_1 & \cdots & \mathbf{a}_m^\top \mathbf{b}_p \end{bmatrix}$$

Matrix-Matrix Multiplications

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} = \mathbb{R}^{n \times p}$.
- Performing $m \times p$ inner products
- or performing n outer products

$$\mathbf{AB} = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & \mathbf{b}_1 & - \\ & \cdots & \\ - & \mathbf{b}_n & - \end{bmatrix} = \mathbf{a}_1\mathbf{b}_1^\top + \cdots + \mathbf{a}_n\mathbf{b}_n^\top$$

Matrix-Matrix Multiplications

- Equivalently written as matrix-vector muliplications

$$\mathbf{AB} = \begin{bmatrix} | & & | \\ \mathbf{Ab}_1 & \cdots & \mathbf{Ab}_p \\ | & & | \end{bmatrix} = \begin{bmatrix} - & \mathbf{a}_1^\top \mathbf{B} & - \\ & \cdots & \\ - & \mathbf{a}_m^\top \mathbf{B} & - \end{bmatrix}$$

Quiz # 3

To multiply $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} = \mathbb{R}^{n \times p}$,
how many scalar multiplications do we need?

Answer

To multiply $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} = \mathbb{R}^{n \times p}$,
how many scalar multiplications do we need?

Answer. $m \times n \times p$.

Because we do $m \times p$ inner prods,

and each inner prod requires n multiplications.

Norms

Norm

- A measure of "length": $\|\cdot\|$ $(: \mathbb{R}^n \to \mathbb{R})$
- Defined by the following properties:
  - Nonnegative: $\|\mathbf{x}\| \geq 0$
  - Definite: $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = \mathbf{0}$
  - Absolute homogeneity: $\|c\mathbf{x}\| = |c| \cdot \|\mathbf{x}\|$
  - Triangle inequality: $\|\mathbf{x}\| + \|\mathbf{y}\| \geq \|\mathbf{x} + \mathbf{y}\|$

Norm

- For a vector $\mathbf{x} \in \mathbb{R}^n$:
  - The ℓ_2 norm: $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$
  - The ℓ_1 norm: $\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_n|$
  - The ℓ_p norm: $\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$
  - The ℓ_∞ norm: $\|\mathbf{x}\|_\infty = \max_{i \in \{1,\ldots,n\}} |x_i|$

Column/Row/Null Space

Linear Independence

- **Linear combination.**

$$\lambda_1 \mathbf{x}_1 + \cdots + \lambda_k \mathbf{x}_k$$

- The vectors $\mathbf{x}_1, \cdots, \mathbf{x}_k$ are linearly independent whenever

$$\lambda_1 \mathbf{x}_1 + \cdots + \lambda_k \mathbf{x}_k = 0 \quad \text{iff} \quad \lambda_1 = \cdots = \lambda_k = 0$$

  - i.e., no vector is a linear combination of remainders.

Span

- The set (space) of all linear combinations

$$\text{span}(\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}) = \left\{ \lambda_1 \mathbf{x}_1 + \cdots + \lambda_k \mathbf{x}_k \;\middle|\; \lambda_i \in \mathbb{R}, \quad \forall i \in [n] \right\}$$

  - *example*. $\mathbb{R}^2$ is spanned by

$$\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

Basis

- A minimal set $A = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ that spans the vector space V.
  - *example*. (one of the) the bases of $\mathbb{R}^2$ is

$$\left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \end{bmatrix} \right\}$$

  - *property*. basis is linearly independent, and adding any element breaks the independence.

Column space

- The column space of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the space spanned by column vectors of $\mathbf{A}$:

$$C(\mathbf{A}) = \left\{ \lambda_1 \mathbf{a}_1 + \cdots + \lambda_n \mathbf{a}_n \;\middle|\; \lambda_i \in \mathbb{R}, \forall i \in [n] \right\} \quad \subseteq \mathbb{R}^m$$

recall p.15

$$\mathbf{W}\mathbf{x} = \begin{bmatrix} | & & | \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ | & & | \end{bmatrix} \mathbf{x} \; = x_1\mathbf{w}_1 + \cdots + x_n\mathbf{w}_n$$

- One can also write:

$$C(\mathbf{A}) = \{ \mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n \}$$

physical meaning: the set of outputs you can get from a model

Row space

- Similarly, the row space is:

$$R(\mathbf{A}) = \{\mathbf{A}^\top \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\} \quad \subseteq \mathbb{R}^n$$

*unfortunately, no clean “physical meaning” as column space...
except that one-to-one correspondence holds between $R(\mathbf{A})$ and $\mathbf{C}(\mathbf{A})$

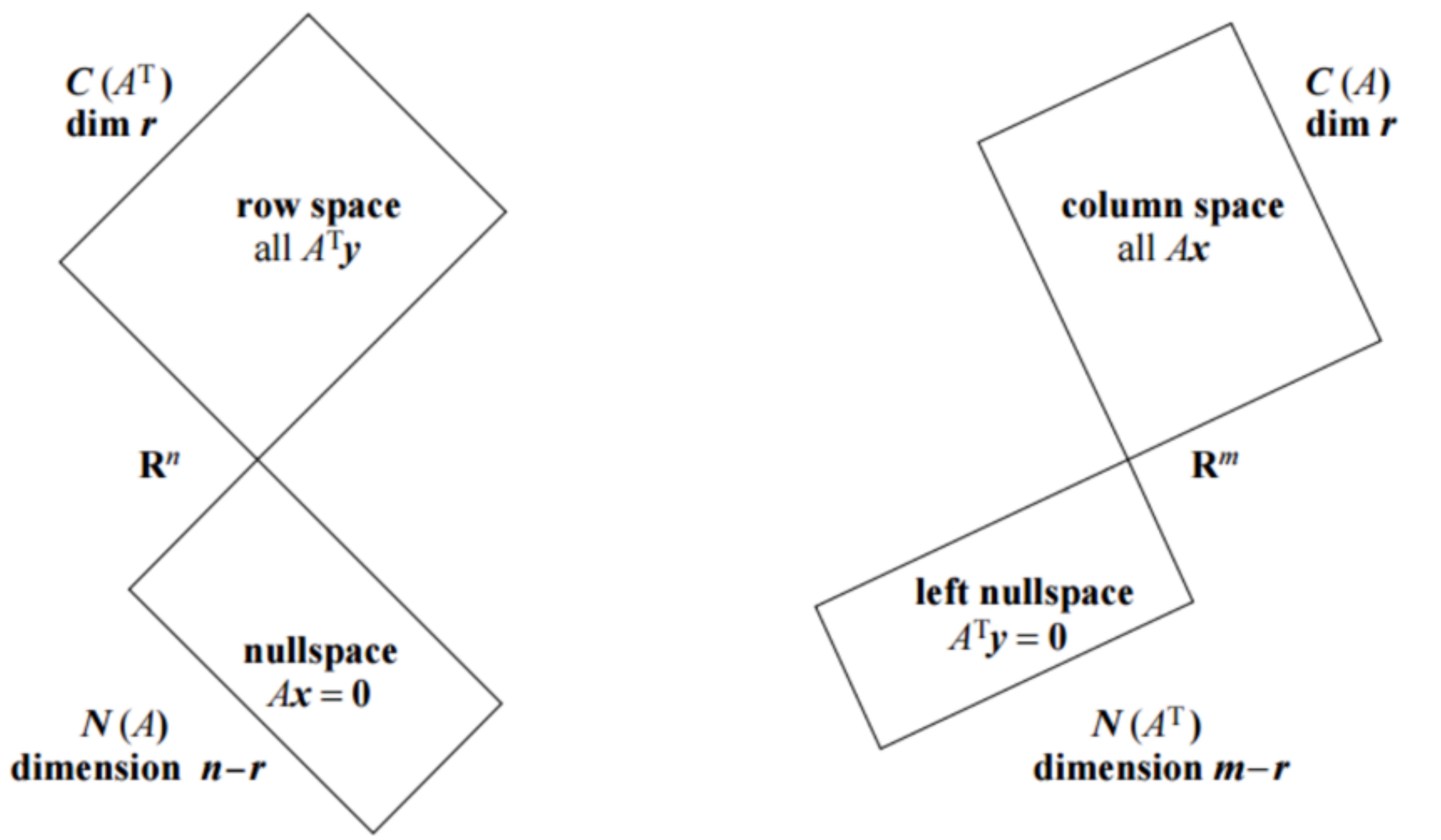
Null space

- The null space of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is

$$N(\mathbf{A}) = \left\{ \mathbf{x} \;\middle|\; \mathbf{A}\mathbf{x} = \mathbf{0}, \quad \mathbf{x} \in \mathbb{R}^n \right\}$$

physical meaning: the set of inputs that you get $\mathbf{0}$ as an output

- The left null space is defined as $N(\mathbf{A}^\top)$ $\in \mathbb{R}^m$



$C(A^T)$
dim r
row space
all A^Ty
R^n
nullspace
$Ax = 0$
$N(A)$
dimension $n-r$
$C(A)$
dim r
column space
all Ax
R^m
left nullspace
$A^Ty = 0$
$N(A^T)$
dimension $m-r$

Rank

- The rank of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is…
  - the number of linearly independent columns
  - the number of linearly independent rows
- *Properties.*
  - $\text{rank}(\mathbf{A}) \leq \min\{m, n\}$
  - $\text{rank}(\mathbf{AB}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$
  - $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$

Inverse

- For a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, the inverse matrix $\mathbf{A}^{-1} \in \mathbb{R}^{n \times n}$ is a matrix such that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$$

(not always invertible—called singular matrix)

- Properties.
  - The inverse exists iff $\mathrm{rank}(\mathbf{A}) = n$ (call this "non-singular")
  - $(\mathbf{A}^{-1})^{-1} = \mathbf{A}, \quad (\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}, \quad (\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$

Special Matrices

Identity Matrix

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ & & \cdots & & \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

- Acts as “1” in the space of matrices

$$\mathbf{A}\mathbf{I}_n = \mathbf{I}_m\mathbf{A} = \mathbf{A}$$

(the system where the input is equal to the output)

Diagonal Matrix

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 & 0 \\ 0 & d_2 & \cdots & 0 & 0 \\ & & \cdots & & \\ 0 & 0 & \cdots & d_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & d_n \end{bmatrix}$$

(the system where each output is a scaled version of input)

Orthogonal/Orthonormal Matrix

- A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is orthogonal when the columns are orthogonal to each other, i.e.,

$$\mathbf{a}_i^\top \mathbf{a}_j = 0, \quad \forall i \neq j$$

- Orthonormal when we further have

$$\|\mathbf{a}_i\|_2 = 1, \quad \forall i \in [n]$$

  - Then, we have $\mathbf{A}^\top \mathbf{A} = \mathbf{A}\mathbf{A}^\top = \mathbf{I}_n$

Property of an orthonormal matrix

- If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is orthonormal,
  - $\mathbf{A}^\top \mathbf{A} = \mathbf{A}\mathbf{A}^\top = \mathbf{I}_n$
  - The matrix preserves the norm, i.e., $\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$.

    Proof. We proceed as

$$\|\mathbf{A}\mathbf{x}\|_2 = \sqrt{(\mathbf{A}\mathbf{x})^\top \mathbf{A}\mathbf{x}} = \sqrt{\mathbf{x}^\top \mathbf{A}^\top \mathbf{A}\mathbf{x}} = \sqrt{\mathbf{x}^\top \mathbf{x}} = \|\mathbf{x}\|_2$$

Symmetric Matrix

- A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric if

$$\mathbf{A}^\top = \mathbf{A}$$

- *Properties*. Real-valued symmetric matrices have
  - real eigenvalues
  - orthogonal eigenvectors
    (useful for SVD)

Definite Matrix

- Positive-semidefinite. For any $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0.$
- *Positive-definite*. For any $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0.$

(similar for negative)

Eigenvalues / Eigenvectors

Eigenvalues & Eigenvectors

- A non-zero vector $\mathbf{x} \in \mathbb{R}^n$ is an eigenvector of $\mathbf{A} \in \mathbb{R}^{n \times n}$ when

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

holds for some λ (the eigenvalue).

physical meaning. output is the same direction as input

- **Determinant** $|\mathbf{A}|$. Product of all eigenvalues.
- **Trace** $\mathrm{Tr}(\mathbf{A})$. Sum of all eigenvalues.

Eigen-decomposition

- Build a column matrix of all (unit norm) eigenvectors, $\mathbf{X}$
  (and Λ a diagonal matrix of respective eigenvalues)
- Then, we have

$$\mathbf{AX} = \mathbf{X}\Lambda.$$

- Sometimes, $\mathbf{X}$ is invertible (diagonalizable) and we can do

$$\mathbf{A} = \mathbf{X}\Lambda\mathbf{X}^{-1}.$$

Eigen-decomposition

$$\mathbf{A} = \mathbf{X}\Lambda\mathbf{X}^{-1}$$

- When this is possible, the "model" $\mathbf{A}$ is sequentially performing:
  - $\mathbf{X}^{-1}$ = send input to another space.
  - Λ = do entrywise scaling
  - $\mathbf{X}$ = pull back to original space.
- *Homework*. Watch https://www.3blue1brown.com/lessons/eigenvalues for visual insights.

Singular Value Decomposition

- SVD decomposes a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ into

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$$

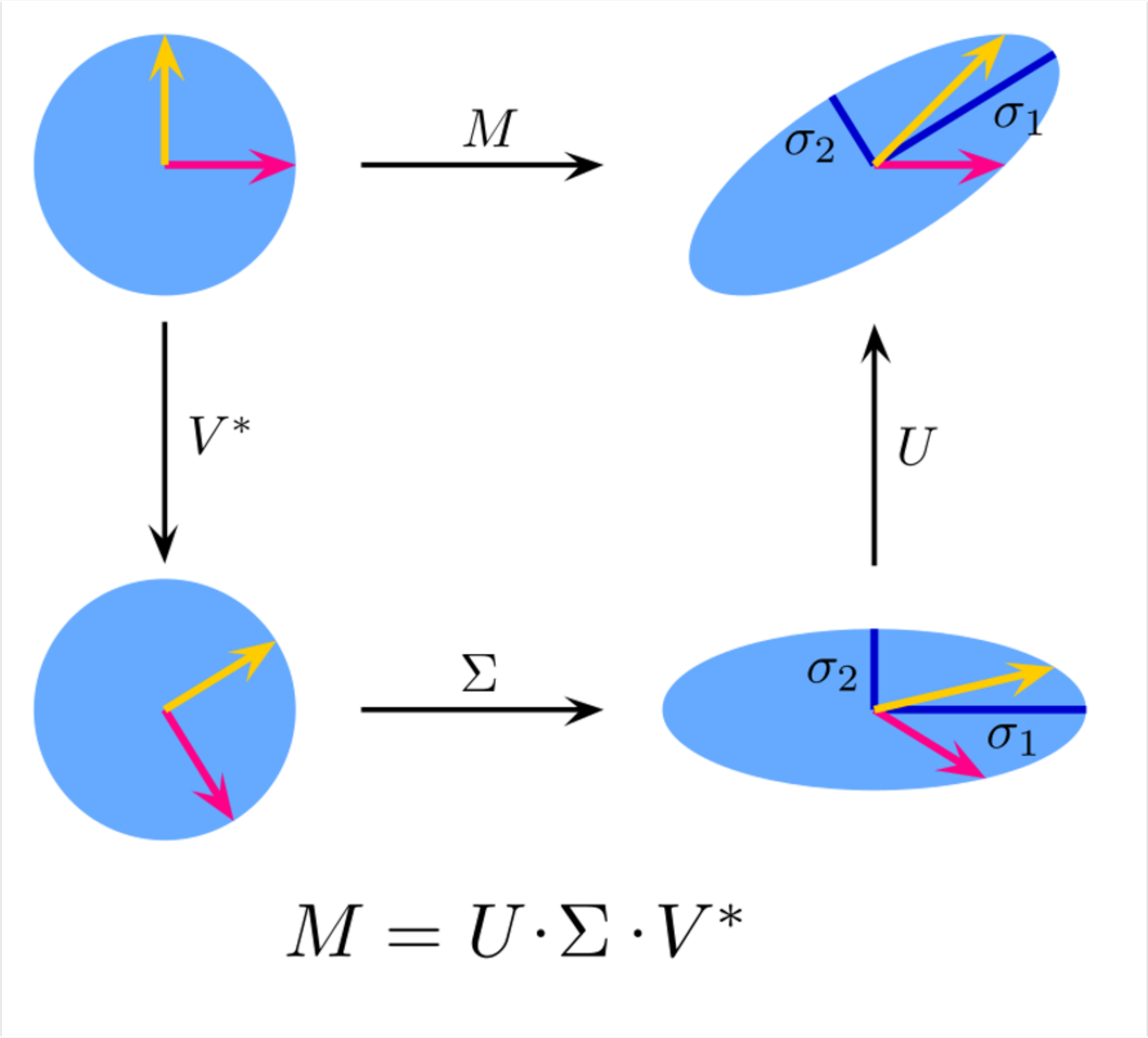
  - $\mathbf{U} \in \mathbb{R}^{m \times m}$ with $\mathbf{U}^\top\mathbf{U} = \mathbf{U}\mathbf{U}^\top = \mathbf{I}_m$
  - $\mathbf{V} \in \mathbb{R}^{n \times n}$ with $\mathbf{V}^\top\mathbf{V} = \mathbf{V}\mathbf{V}^\top = \mathbf{I}_n$
  - $\mathbf{\Sigma}$ is a diagonal matrix (with zero paddings).

Singular Value Decomposition

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$$

- **How?**
  - Construct $\mathbf{U}$ with eigenvectors of $\mathbf{A}\mathbf{A}^\top$.
    - $\mathbf{A}\mathbf{A}^\top$ is real symmetric, and thus have orthogonal eigenvectors.
  - Construct $\mathbf{V}$ with eigenvectors of $\mathbf{A}^\top\mathbf{A}$.
  - Compute $\mathbf{\Sigma}$ with the square-root of eigenvalues of $\mathbf{A}^\top\mathbf{A}$.

Singular Value Decomposition



Cheers

- *Next up.* Gram-Schmidt, Matrix Calculus, Basic Probability.