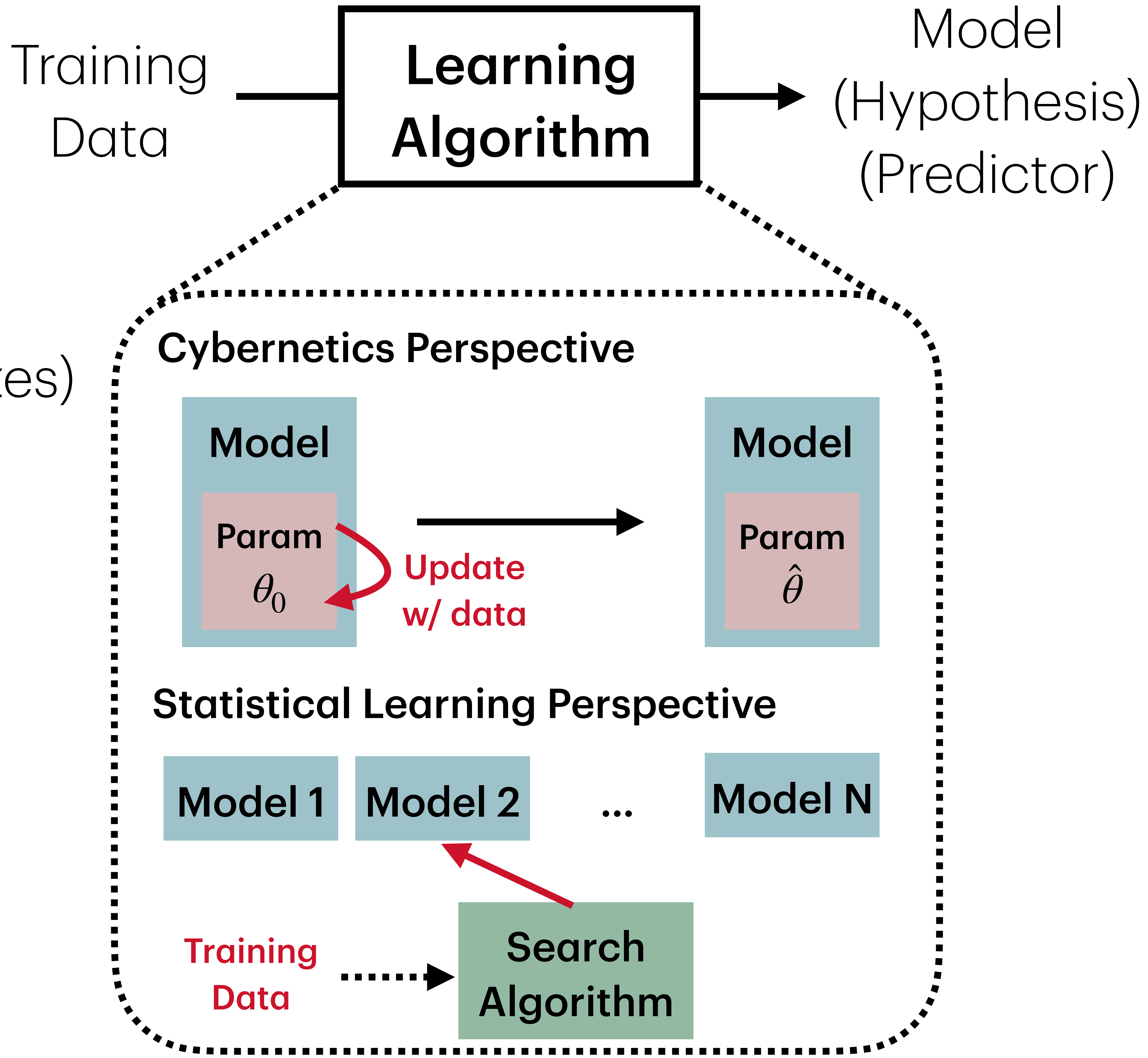
Recap: Linear Algebra

EECE454 Intro. to Machine Learning Systems

Fall 2024

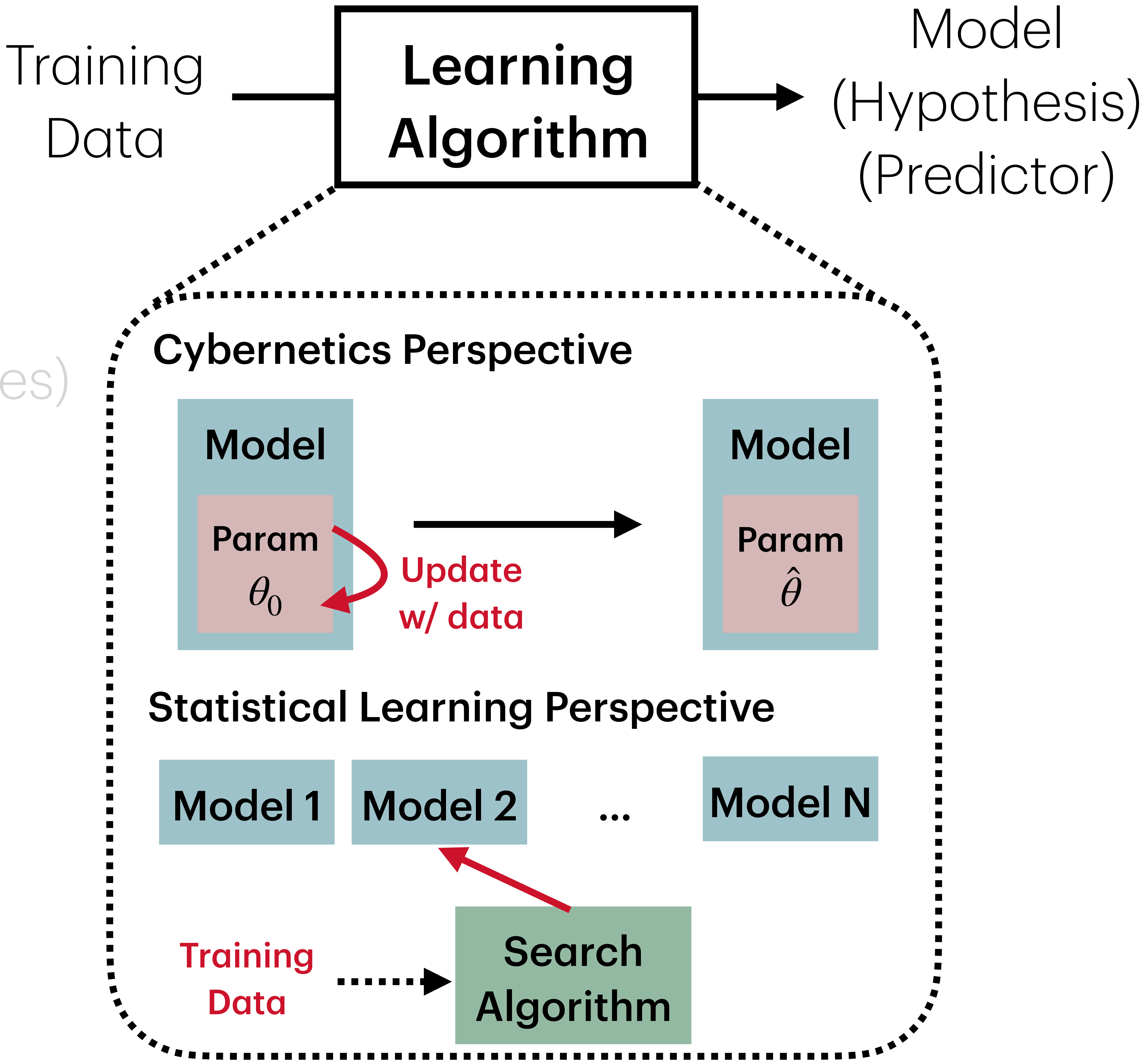
Last class

- An extremely rough description about ML
  - We have some many models at hand
    - potentially parametrized by some state θ
  - ML algorithm selects the right model (i.e., optimizes) by evaluating the model on the data
  - The selected model is deployed to new data



Last class

- An extremely rough description about ML
  - We have some many models at hand
    - potentially parametrized by some state θ
  - ML algorithm selects the right model (i.e., optimizes) by evaluating the model on the data
  - The selected model is deployed to new data
- Did not talk about:
  - How to formalize the models, how to optimize, how to capture the randomness of the data



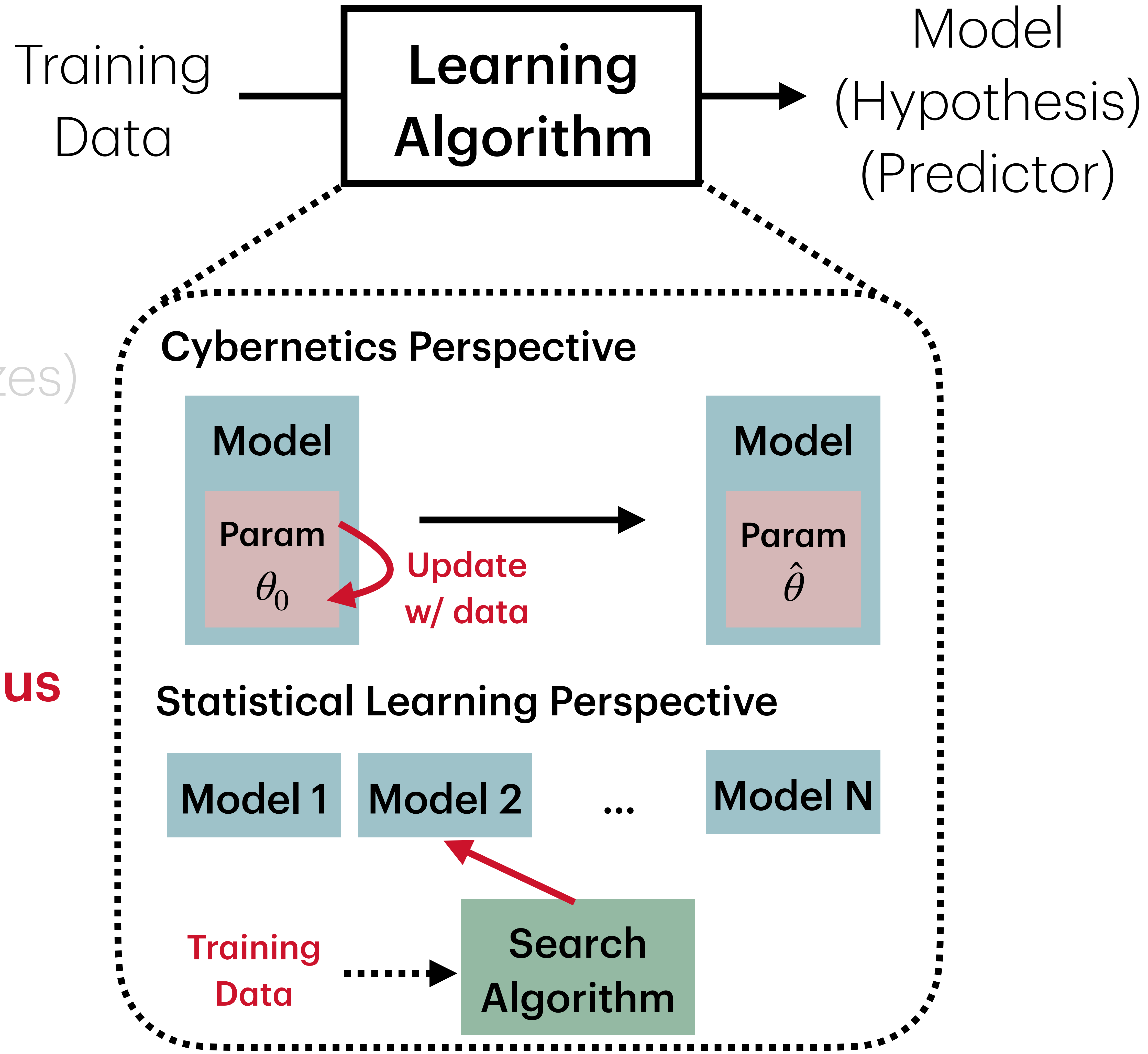
Last class

- An extremely rough description about ML
  - We have some many models at hand
    - potentially parametrized by some state θ
  - ML algorithm selects the right model (i.e., optimizes) by evaluating the model on the data
  - The selected model is deployed to new data
- Did not talk about:
  - How to formalize the models, how to optimize, how to capture the randomness of the data

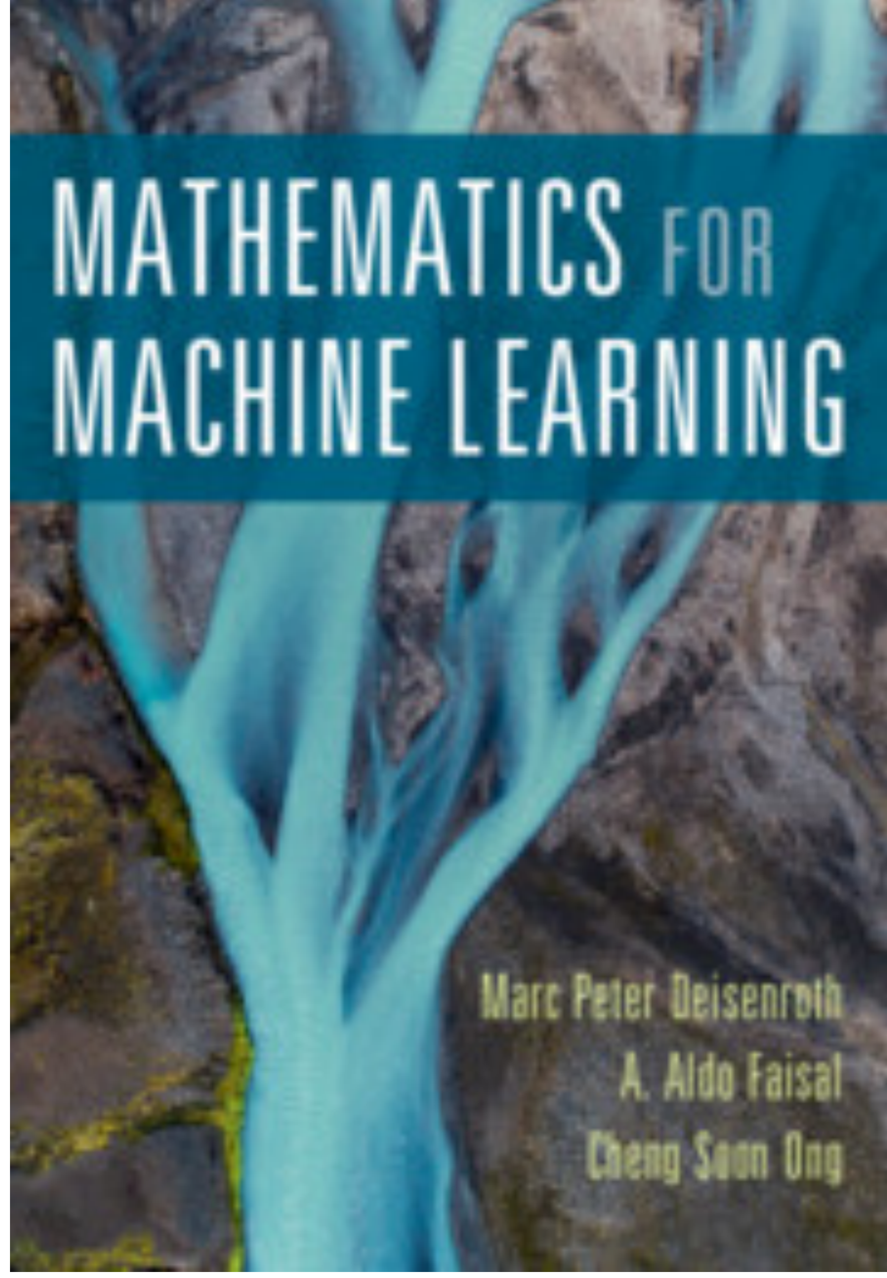
Vectors & Matrices

Matrix Calculus

Probability

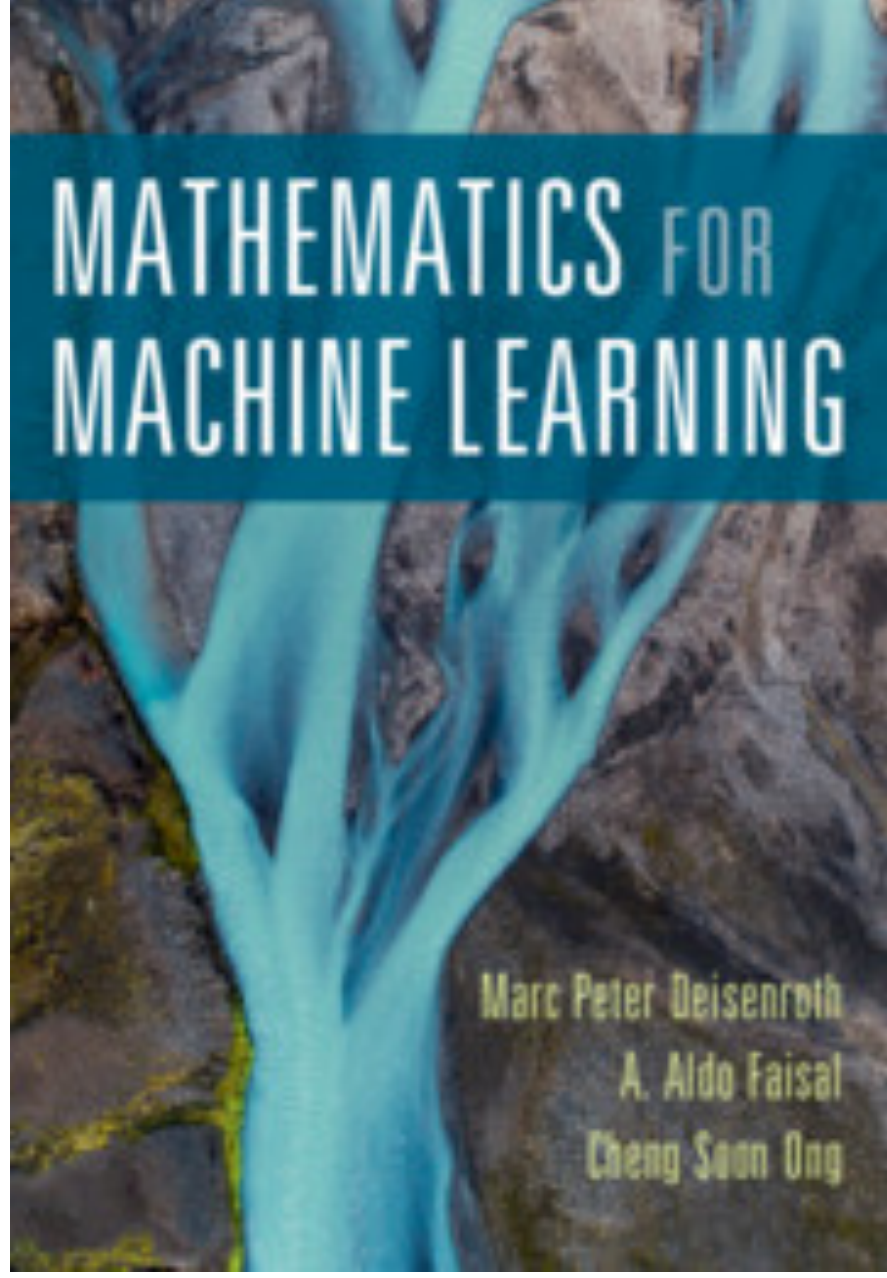


Today



- Brief recap of linear algebra
  - MML book: Chapter 1 — Chapter 6
  - D2DL: Section 2.3. — 2.6.
  - https://cs229.stanford.edu/lectures-spring2022/cs229-linear_algebra_review.pdf
  - https://www.3blue1brown.com/topics/linear-algebra
- **Next week.** Probability & optimization

Today



- Brief recap of linear algebra
  - MML book: Chapter 1 — Chapter 6
  - D2DL: Section 2.3. — 2.6.
  - https://cs229.stanford.edu/lectures-spring2022/cs229-linear_algebra_review.pdf
  - https://www.3blue1brown.com/topics/linear-algebra
- **Next week.** Probability & optimization

- **Disclaimer.** Boring & incomplete; use the slides as a **guide for self-study**

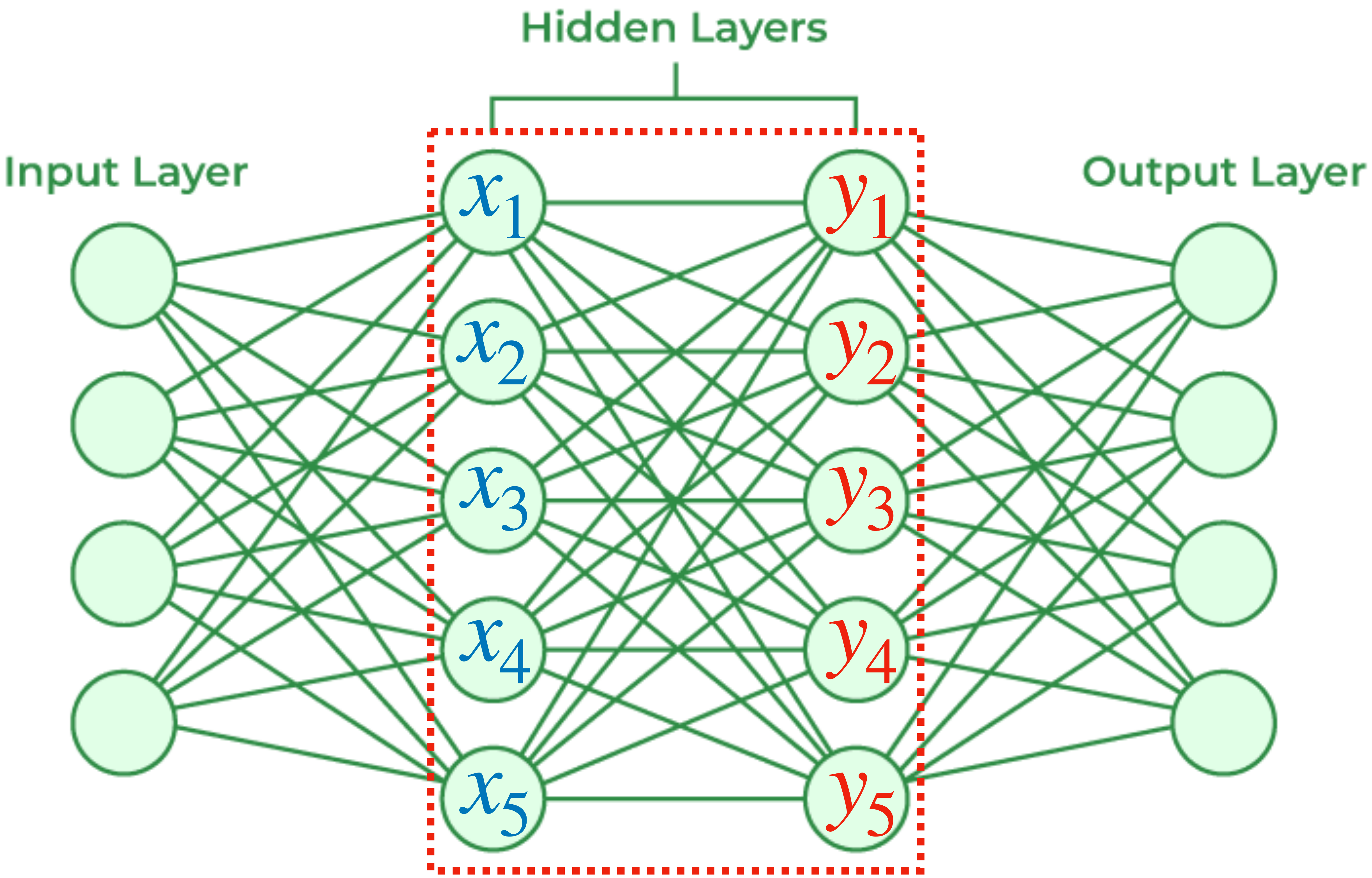
Experimental feature

- For the sake of not-being-boring, let us go through this session with **Quiz-like format**.
  - Login to slido.com with your mobile
    - Enter the code #1794667
  - Alternatively, use the QR code

Why matrices?

Why matrices?

- Matrices are the **simplest model** of the relationship between multidimensional input & output.
  - Used as a building block of more elaborate systems, e.g., neural nets.
  - Used for characterizing **models**, **data**, ...



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

Formalisms

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

Formalisms

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 $\mathcal{B}$

Formalisms

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 $\mathbf{x} \in \mathbb{R}^n$ (bold lowercase)
  This is ...

(a)

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

(b)

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

Quiz #1

- Let there be a vector $\mathbf{x} \in \mathbb{R}^n$ (bold lowercase)
  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)

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

Quiz #2

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

$m \times n$ means 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

Products of vectors

- There are two different types: Inner, and Outer

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

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

- $\mathbf{x}$ and $\mathbf{y}$ to have same dimensions
- You will use it on a daily basis
- Alternate notation: $\langle \mathbf{x}, \mathbf{y} \rangle$
- Intuition: alignedness of vectors

Outer product

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

- Can have different dimensions
- Intuition: measuring alignedness of each entries (scalar)

Matrix-Vector Multiplication

- Performing multiple inner products with row vectors
    - Measuring alignedness of input with m reference vectors (some sort of dictionaries)

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

Matrix-Vector Multiplication

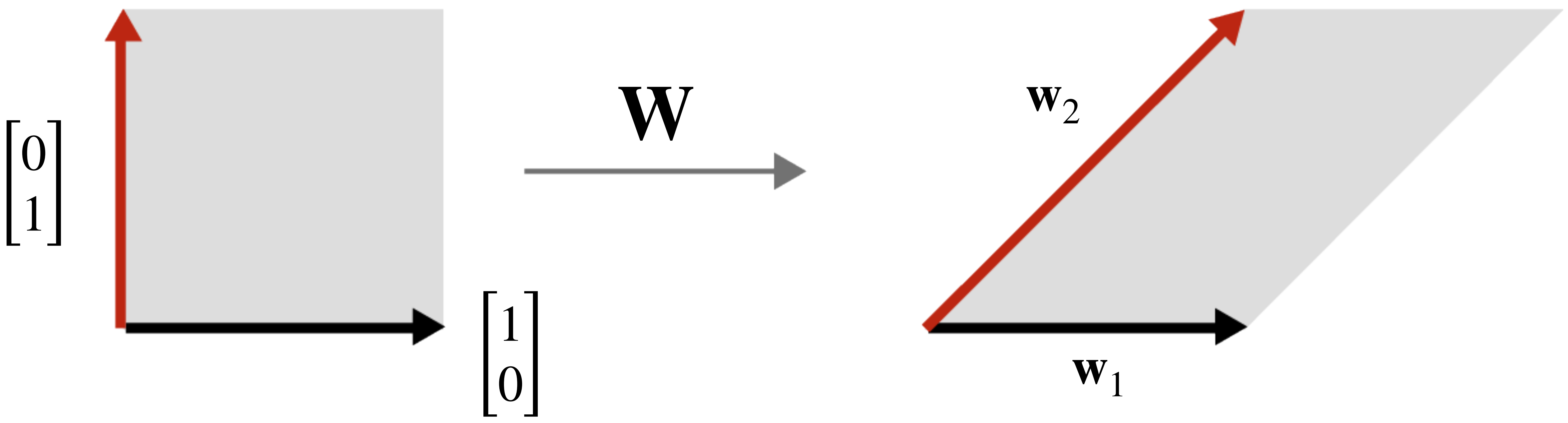
- Performing multiple inner products with row vectors
  - Measuring alignedness of input with m different internal states (some sort of dictionaries)
- Alternatively, we are taking a weighted sum of column vectors
  - Inputs are recipes, columns are ingredients, and output is the food.

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

MVM: System perspective

- The matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be viewed as an 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 Multiplication

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} = \mathbb{R}^{n \times p}$
- Performing a **matmul** is equivalent to performing mp inner products
  - measuring alignedness between m reference vectors and p input signals

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

Matrix-Matrix Multiplication

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} = \mathbb{R}^{n \times p}$
- Performing a **matmul** is equivalent to performing mp inner products
  - measuring alignedness between m reference vectors and p input signals
- Alternatively, 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 Multiplication

- Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} = \mathbb{R}^{n\times p}$
- Performing a **matmul** is equivalent to performing mp inner products
    - measuring alignedness between m reference vectors and p input signals
- Alternatively, performing n outer products
- Or p (or m) matrix-vector multiplications

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

Quiz #3

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

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

Quiz #3

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

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

- **Answer.** mnp
  - We do mp inner products
  - Each inner product requires n multiplications.

Norms

Norm

- A measure of length: $\|\cdot\|$
  - A function $\mathbb{R}^n \rightarrow \mathbb{R}$
- Defined axiomatically by the following properties:
  - Nonnegativity: $\|\mathbf{x}\| \geq 0$
  - Definiteness: $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
  - 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: $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$
    - That is, $\|\mathbf{x}\|_2^2 = \mathbf{x}^\top \mathbf{x}$
  - The ℓ_1 norm: $\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_n|$
  - The ℓ_p norm: $\|\mathbf{x}\|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}$
  - The ℓ_∞ norm: $\|\mathbf{x}\|_\infty = \max_{i \in \{1,\ldots,n\}} |x_i|$

Quiz #4

- Let us define the quantity ℓ_0 norm as $\|\mathbf{x}\|_0 = |x_1|^0 + \cdots + |x_n|^0$
    - Assume that we use the convention $0^0 = 0$
    - That is, the ℓ_0 norm counts the number of nonzeros.
- **Question.** Is this really a norm?

Quiz #4

- Let us define the quantity ℓ_0 norm as $\|\mathbf{x}\|_0 = |x_1|^0 + \cdots + |x_n|^0$
    - Assume that we use the convention $0^0 = 0$
    - That is, the ℓ_0 norm counts the number of nonzeros.
- **Question.** Is this really a norm?
- **Answer.** A formal proof as a homework :P

Column / Row / Null Space

Linear Combination

- The **linear combination** of k different vectors can be written as

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

Linear Combination

- The **linear combination** of k different vectors can be written as

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

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

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

  - That is, no vector is a linear combination of the others.

Linear Combination

- The **linear combination** of k different vectors can be written as

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

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

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

  - That is, no vector is a linear combination of the others.

- The **span** is the set of all linear combinations

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

  - Example. $\mathbb{R}^2$ is spanned by $\{[1,0]^\top, [0,1]^\top\}$

Basis

- The **basis** of a vector space V is a minimal set $A = \{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$ such that

$$\text{span}(A) = V$$

  - Example. One possible choice of the basis of $\mathbb{R}^2$ is

$$\{[1,3]^\top, [4,1]^\top\}$$

  - Property 1. Basis is linearly independent.

  - Property 2. Adding any element to the basis breaks the linear independence.

Column Space

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

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

  - A subspace of $\mathbb{R}^m$

Column Space

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

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

  - A subspace of $\mathbb{R}^m$

  - One can also write

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

recall...

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

  - Physical meaning. The set of outputs you can get from a model $\mathbf{A}$
    - If the column space of your model does not contain the desired prediction outcomes, perhaps you should modify your model.

Row Space

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

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

  - A subspace of $\mathbb{R}^n$
  - No clean physical meaning by itself
    - One-to-one correspondence holds between $R(\mathbf{A})$ and $C(\mathbf{A})$

Null Space

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

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

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

Null Space

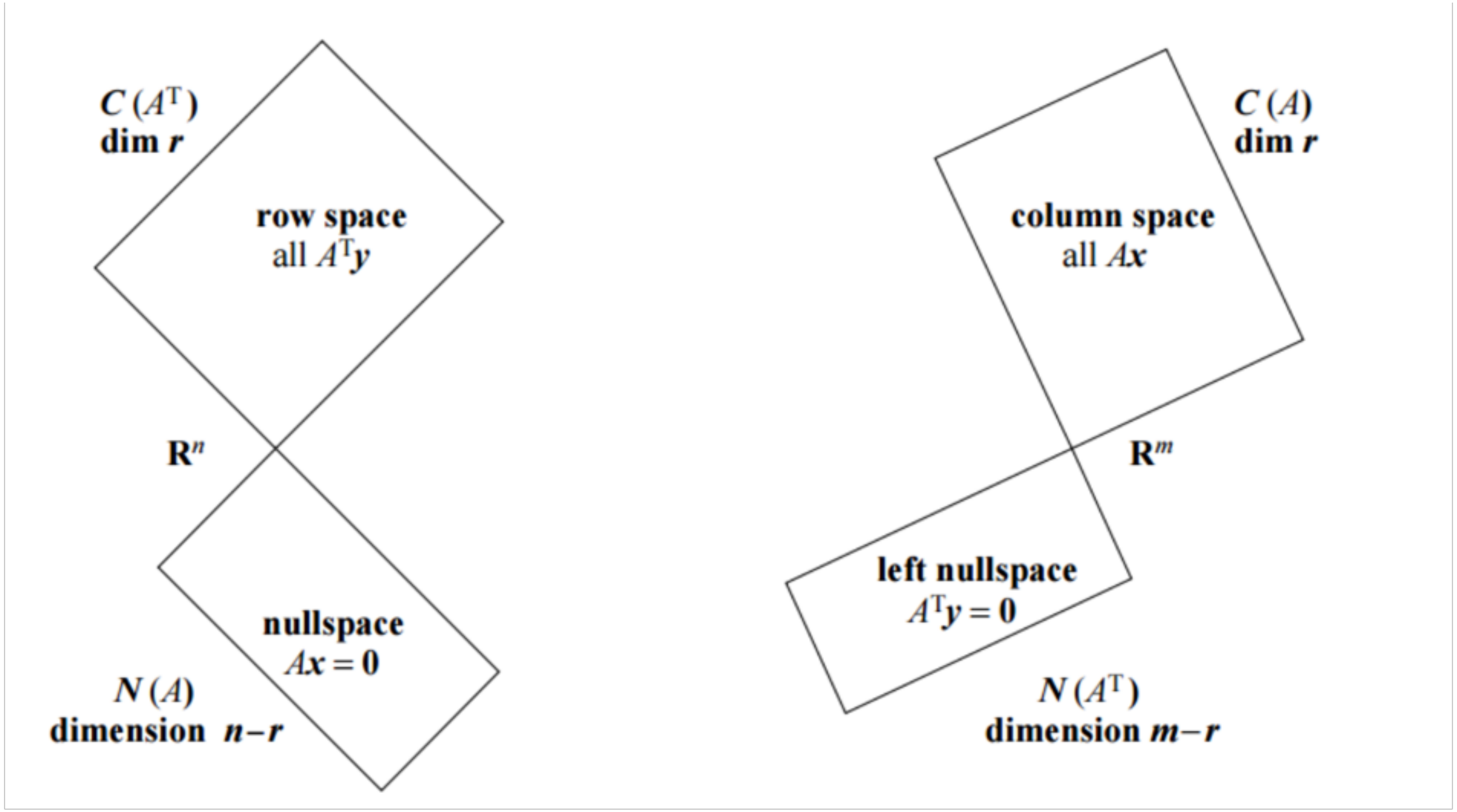
- The **null space** of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as

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

  - A subspace of $\mathbb{R}^n$
  - The (left) null space is defined as $N(\mathbf{A}^\top)$
  - Physical meaning. The set of inputs that you get $\mathbf{0}$ as a prediction
    - If you add null inputs to another input, the outcome will not change
    - The model cannot detect such change (or is robust to).

Null Space

- Property. The row space is an orthogonal complement of the null space.
  - Thus, the row space can be viewed as the vectors the model $\mathbf{A}$ is sensitive to



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.
        - $\mathrm{rank}(\mathbf{A}) \leq \min\{m, n\}$
        - $\mathrm{rank}(\mathbf{AB}) \leq \min\{\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B})\}$
        - $\mathrm{rank}(\mathbf{A} + \mathbf{B}) \leq \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$

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.
        - $\operatorname{rank}(\mathbf{A}) \leq \min\{m, n\}$
        - $\operatorname{rank}(\mathbf{AB}) \leq \min\{\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})\}$
        - $\operatorname{rank}(\mathbf{A} + \mathbf{B}) \leq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B})$
    - Physical meaning. If we have low rank, we can remove dependent rows/columns to make the matrix smaller, reducing computations.

Inverse

- Given 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; non-invertible matrices are called singular.
  - Properties.
    - The inverse exists iff $\mathbf{rank}(\mathbf{A}) = n$
    - $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
    - $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
    - $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$

Special Matrices

Identity Matrix

- The **identity matrix** is defined as a matrix that gives identical output as the input when multiplied

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

  - Same as "1" in the space of matrices.
  - This is simply

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

  - <u>Physical meaning</u>. A system whose input is same as an output
    - desirable property of information transmission, e.g., "cable" or "memory bus"

Diagonal Matrix

- The **diagonal matrix** is a matrix with nonzero elements only on the diagonal, i.e.,

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

- Physical meaning. A model which each output entry is a scaled version of input
    - If this is our predictor, it may require very few computations

Orthogonal / Orthonormal Matrix

- An **orthogonal matrix** $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a matrix whose columns are orthogonal to each other

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

Orthogonal / Orthonormal Matrix

- An **orthogonal matrix** $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a matrix whose columns are orthogonal to each other

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

- An **orthonormal matrix** is an orthogonal matrix with

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

  - Property 1. We have $\mathbf{A}^\top \mathbf{A} = \mathbf{A}\mathbf{A}^\top = \mathbf{I}_n$
  - Property 2. The matrix preserves the norm, i.e., $\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$
    - ***Proof.*** Volunteer?

Symmetric Matrix

- A **symmetric matrix** $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a matrix such that

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

  - Property. Have real eigenvalues and orthogonal eigenvectors (useful for SVD)
  - Examples. Covariance matrices, the matrices generated as $\mathbf{B}\mathbf{B}^\top$

Symmetric Matrix

- A **symmetric matrix** $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a matrix such that

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

  - Property. Have real eigenvalues and orthogonal eigenvectors (useful for SVD)
  - Examples. Covariance matrices, the matrices generated as $\mathbf{B}\mathbf{B}^\top$

- A **positive-definite matrix** $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a matrix such that

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0, \quad \forall \mathbf{x} \neq \mathbf{0}$$

  - Semidefinite if holds with $\geq$ instead of $>$

Eigenvalues and Eigenvectors

Eigenvalues & Eigenvectors

- An **eigenvector** $\mathbf{x} \in \mathbb{R}^n$ of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a nonzero vector such that

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

holds for some λ (called the **eigenvalue**).

Eigenvalues & Eigenvectors

- An **eigenvector** $\mathbf{x} \in \mathbb{R}^n$ of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a nonzero vector such that

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

holds for some λ (called the **eigenvalue**).

  - Physical meaning. The inputs for which the model performs only scaling
    - Useful as a basis
  - Determinant $|\mathbf{A}|$ is the product of all eigenvalues
  - Trace $\mathbf{Tr}(\mathbf{A})$ is the sum of all eigenvalues

Eigen-decomposition

- Suppose that we have a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$

- Consider building a column matrix $\mathbf{U}$ of all (unit norm) eigenvectors of $\mathbf{A}$.

    - Then, we have

$$\mathbf{AU} = \mathbf{U}\mathbf{\Lambda}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of all respective eigenvalues.

Eigen-decomposition

- Suppose that we have a square matrix $\mathbf{A}$
- Consider building a column matrix $\mathbf{U}$ of all (unit norm) eigenvectors of $\mathbf{A}$.
  - Then, we have

$$\mathbf{A}\mathbf{U} = \mathbf{U}\Lambda$$

  where Λ is a diagonal matrix of all respective eigenvalues.

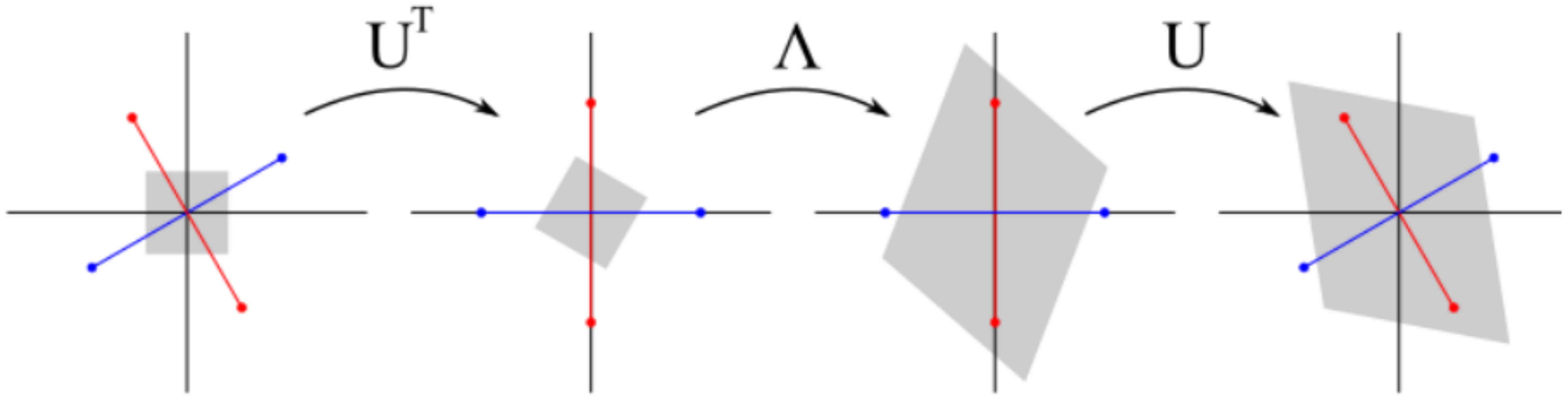
    - If $\mathbf{U}$ is invertible, the matrix $\mathbf{A}$ is said to be **diagonalizable**, and we can write

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^{-1} = \mathbf{U}\Lambda\mathbf{U}^\top$$

Eigen-decomposition

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

- Whenever this is doable, the model $\mathbf{A}$ is actually performing:
  - $\mathbf{U}^\top$: Send input to another space
  - Λ: Perform scaling for each dimension
  - $\mathbf{U}$: Pull back to the original space



- **Further material.** Watch https://www.3blue1brown.com/lessons/eigenvalues for visual insights.

Singular Value Decomposition

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

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

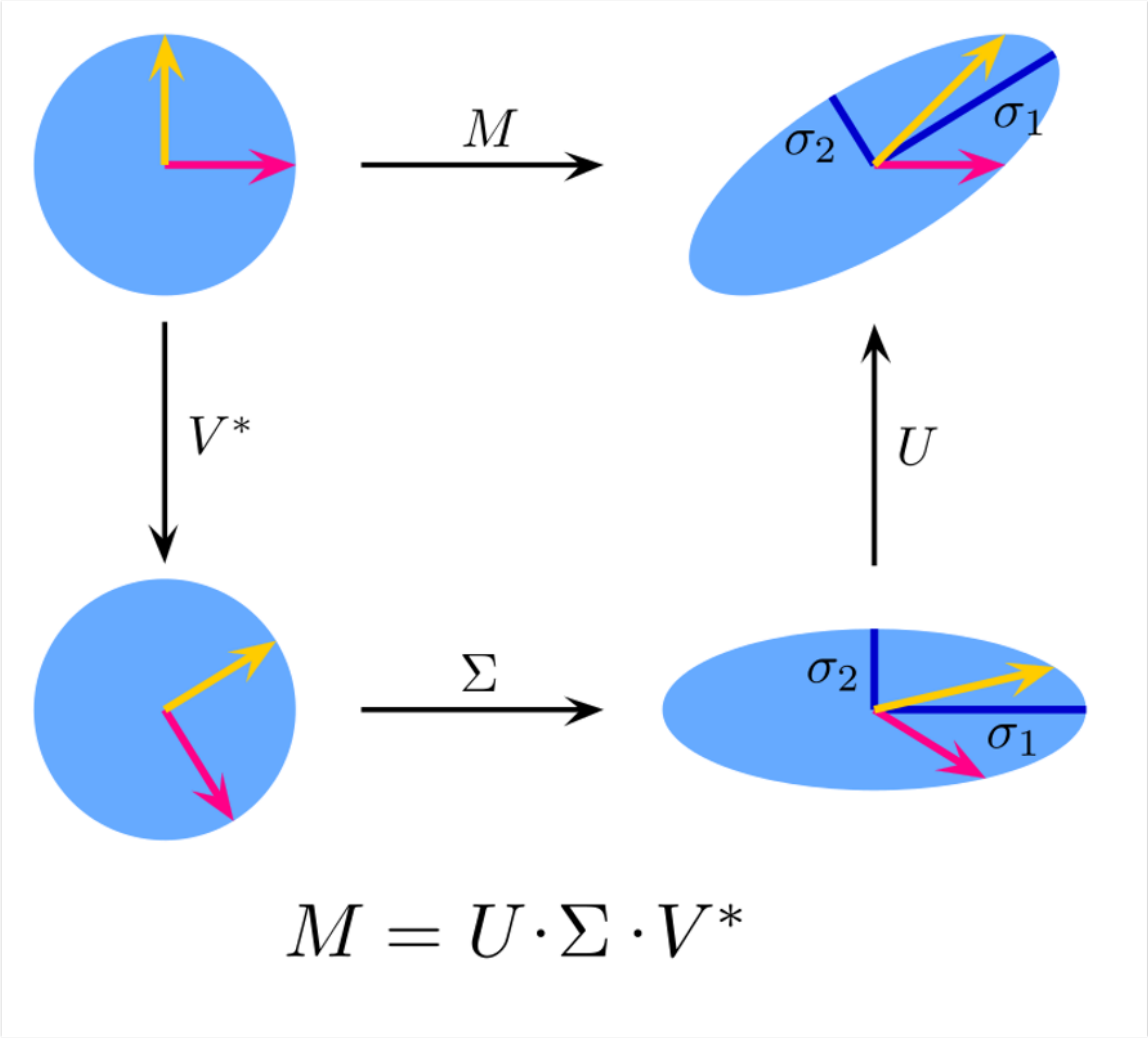
  - $\mathbf{U} \in \mathbb{R}^{m \times m}$ with $\mathbf{U}^\top\mathbf{U} = \mathbf{U}\mathbf{U}^\top = \mathbf{I}_m$
  - $\mathbf{V} \in \mathbb{R}^{n \times n}$ with $\mathbf{V}^\top\mathbf{V} = \mathbf{V}\mathbf{V}^\top = \mathbf{I}_n$
  - $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is a diagonal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

Singular Value Decomposition

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

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

Singular Value Decomposition



Wrapping Up

- **Today.** We have gone through basic linear algebra
- **Next class.** Gram-Schmidt, Matrix Calculus (optimization), Basic Probability

Cheers