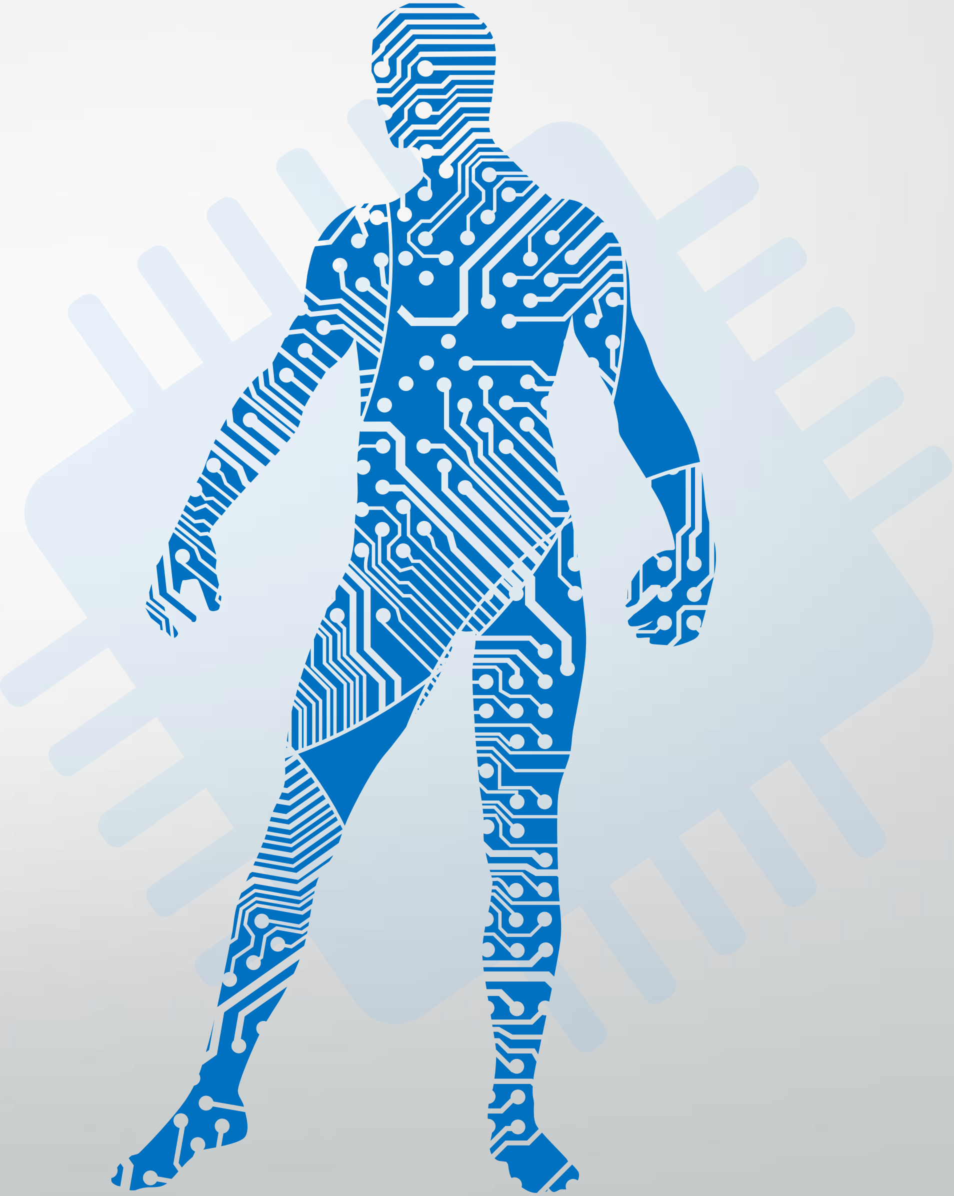
MACHINE LEARNING

LINEAR ALGEBRA REVIEW

AGENDA

01 Linear Algebra Review

Notation, properties, operations, calculus



LINEAR ALGEBRA

NOTATION

MULTIPLICATION

LINEAR ALGEBRA
BASIC NOTATION

Systems of **equations** as matrices:

$$\begin{array}{rcrcr} 4x_1 & - & 5x_2 & = & -13 \\ -2x_1 & + & 3x_2 & = & 9. \end{array}$$

$$Ax = b$$

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}.$$

LINEAR ALGEBRA
BASIC NOTATION

We have a matrix $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{mxn}}$ and a vector $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

LINEAR ALGEBRA
MULTIPLICATION

Multiplication between **2 matrices** $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{mxn}}$ and $\boldsymbol{B} \in \mathbb{R}^{\boldsymbol{nxp}}$ gives as a result a matrix $\boldsymbol{C} \in \mathbb{R}^{\boldsymbol{mxp}}$.

$$C = AB \in \mathbb{R}^{m \times p},$$

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

LINEAR ALGEBRA
MULTIPLICATION

EXAMPLE:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

LINEAR ALGEBRA
MULTIPLICATION

ANSWER:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

$$C = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} & a_{21}b_{13} + a_{22}b_{23} \end{bmatrix}$$

LINEAR ALGEBRA

MULTIPLICATION

Inner product (dot): multiplication between **2 vectors** $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{\boldsymbol{n}}$

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{vmatrix} = \sum_{i=1}^{n} x_i y$$

LINEAR ALGEBRA
MULTIPLICATION

External product: **multiplication** between **2 vectors** $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{m}}, \boldsymbol{y} \in \mathbb{R}^{\boldsymbol{n}}$

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

LINEAR ALGEBRA
MULTIPLICATION

Multiplication between **vector** $x \in \mathbb{R}^n$ and **matrix** $A \in \mathbb{R}^{mxn}$.

$$y = Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \\ a^1 \\ \\ \end{bmatrix} x_1 + \begin{bmatrix} \\ a^2 \\ \\ \end{bmatrix} x_2 + \ldots + \begin{bmatrix} \\ a^n \\ \\ \end{bmatrix} x_n$$

$$\begin{aligned} y^T &= x^T A \\ &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} \\ &= x_1 \begin{bmatrix} - & a_1^T & - \end{bmatrix} + x_2 \begin{bmatrix} - & a_2^T & - \end{bmatrix} + \ldots + x_n \begin{bmatrix} - & a_n^T & - \end{bmatrix} \end{aligned}$$

LINEAR COMBINATION

LINEAR ALGEBRA
MULTIPLICATION PROPERTIES

ASSOCIATIVITY

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$

NONCOMMUTATIVE

$$\boldsymbol{AB} \neq \boldsymbol{BA}$$

DISTRIBUTIVITY

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC})$$

LINEAR ALGEBRA

OPERATIONS

PROPERTIES

LINEAR ALGEBRA

IDENTITY MATRIX

The **identity matrix,** $I \in \mathbb{R}^{nxn}$, is a **square matrix** that has **ones** in its **diagonal** and **zeros elsewhere** .

Given a matrix $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{mxn}}$:

$$AI = A = IA$$

LINEAR ALGEBRA

T H E   T R A N S P O S E

The **transpose** of a **matrix** results by "**flipping**" the **rows** and **columns**.

Given a **matrix** $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{nxm}}$, the **transposed elements** would be **defined** by:

$$(A^T)_{ij} = A_{ji}.$$

Properties:

$$(A^T)^T = A$$

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

$$(A+B)^T = A^T + B^T$$

LINEAR ALGEBRA
SYMMETRIC MATRICES

A **square matrix** $A \in \mathbb{R}^{nxn}$ is **symmetric** if:

$$A = A^T$$

A **square matrix** $A \in \mathbb{R}^{nxn}$ is **anti- symmetric** if:

$$A = -A^T$$

LINEAR ALGEBRA

THE TRACE

The trace of a **square matrix** $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{nxn}}$, denoted by $\boldsymbol{tr(A)}$**,** is the **sum** of the **diagonal elements** of the matrix.

$$\mathrm{tr}A = \sum_{i=1}^{n} A_{ii}.$$

Trace properties given $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{\boldsymbol{nxn}}$ and $\boldsymbol{t} \in \mathbb{R}$:

$$\boldsymbol{trA = trA^T}$$

$$\boldsymbol{tr(A+B) = trA + trB}$$

$$\boldsymbol{tr(tA) = t\ tra(A)}$$

$$\boldsymbol{tr(AB) = tr(BA)}$$

LINEAR ALGEBRA

VECTOR NORMS

The **norm** of a **vector** (**informally**) is a **measure** of the "***length***" of the **vector**.

EXAMPLE: the **L2 norm** of a vector has the following expression.

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Formally, the **norm** of a **vector** is **any function** $\boldsymbol{f}: \mathbb{R}^{\boldsymbol{n}} \rightarrow \mathbb{R}$ that **satisfies 4 properties**:

$$\forall \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}, \boldsymbol{f}(\boldsymbol{x}) \geq \mathbf{0}$$
$$\boldsymbol{f}(\boldsymbol{x}) = \mathbf{0} \leftrightarrow \boldsymbol{x} = \mathbf{0}$$
$$\forall \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}, \boldsymbol{t} \in \mathbb{R}, \boldsymbol{f}(\boldsymbol{t}\boldsymbol{x}) = |\boldsymbol{t}|\boldsymbol{f}(\boldsymbol{x})$$
$$\boldsymbol{f}(\boldsymbol{x} + \boldsymbol{y}) \leq \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{y})$$

LINEAR ALGEBRA

LINEAR INDEPENDENCE

A **set** of **vectors** $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}\} \subset \mathbb{R}^{\boldsymbol{m}}$ are **LINEARLY INDEPENDENT** if **no vector** can be represented as a **linear combination** of the remaining vectors.

If a vector from the **vector set** $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}\} \subset \mathbb{R}^{\boldsymbol{m}}$ can be represented as a **linear combination** of the remaining vectors, then the vectors are said to be **LINEARLY DEPENDENT**.

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

Rango de una matriz: es el tamaño del conjunto más grande de **columnas** de **A** que comprenden un **conjunto** de **vectores linealmente independientes**.

LINEAR ALGEBRA

R A N K

The **rank** of a **matrix** is the **size** of the **largest subset** of **columns** of $\boldsymbol{A}$ (**column rank**) or **rows** of $\boldsymbol{A}$ (**row rank**) that constitute a **linearly independent set** of **vectors**.

For any matrix $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{m} \times \boldsymbol{n}}$, the **column rank** of $\boldsymbol{A}$ is **equal** to the **row rank** of $\boldsymbol{A}$**.**

Properties:

- For $A \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A) \leq \min(m, n)$. If $\operatorname{rank}(A) = \min(m, n)$, then A is said to be ***full rank***.
- For $A \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.
- For $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $\operatorname{rank}(AB) \leq \min(\operatorname{rank}(A), \operatorname{rank}(B))$.
- For $A, B \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A + B) \leq \operatorname{rank}(A) + \operatorname{rank}(B)$.

LINEAR ALGEBRA

D E T E R M I N A N T

The **determinant** of a **square matrix** $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{nxn}}$ is a **function** $\mathbb{R}^{\boldsymbol{nxn}} \rightarrow \mathbb{R}$ denoted by $\boldsymbol{detA}$ or $|\boldsymbol{A}|$.

Given a **set of points** $\boldsymbol{S} \subset \mathbb{R}^{\boldsymbol{n}}$ generated by **taking all possible linear combinations** of the **row vectors** $\boldsymbol{a_1, \dots, a_n} \in \mathbb{R}^{\boldsymbol{n}}$ of $\boldsymbol{A}$, where the **coefficients** of the **linear combination** $\boldsymbol{\alpha_1, \alpha_2, \dots, \alpha_n}$ **satisfy** $\boldsymbol{0 \leq \alpha_i \leq 1, i = 1, \dots, n}$.

$$S = \{v \in \mathbb{R}^n : v = \sum_{i=1}^{n} \alpha_i a_i \text{ where } 0 \leq \alpha_i \leq 1, i = 1, \dots, n\}.$$

The **absolute value** of the **determinant** is a **measure** of the $\boldsymbol{n}$ **dimensional *"volume"*** of the **parallelotope** formed by $\boldsymbol{S^n}$.

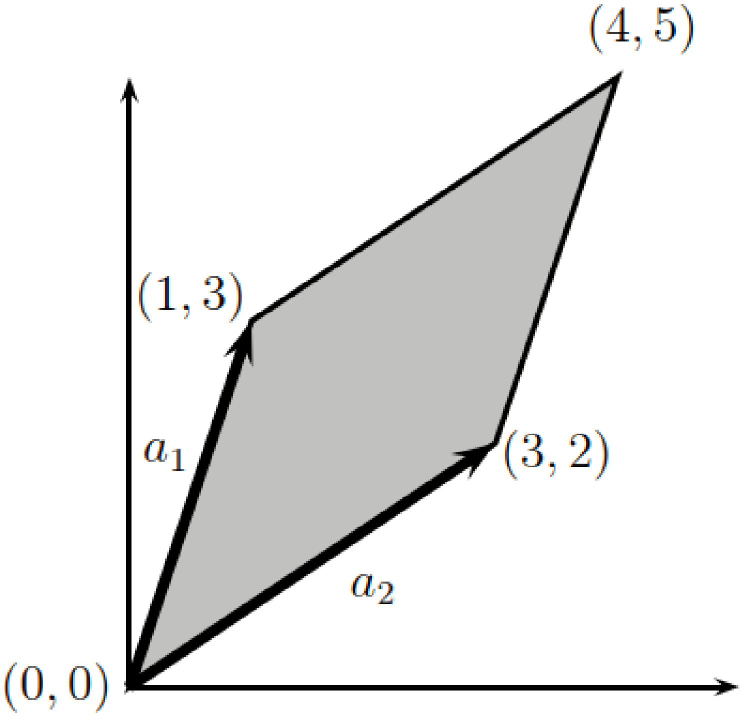
LINEAR ALGEBRA

DETERMINANT

EXAMPLE:

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}.$$

$$a_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad a_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$



LINEAR ALGEBRA

D E T E R M I N A N T

The **general recursive** formula for the **determinant** is:

$$
\begin{aligned}
|A| &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{\backslash i, \backslash j}| \qquad (\text{for any } j \in 1, \ldots, n) \\
&= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{\backslash i, \backslash j}| \qquad (\text{for any } i \in 1, \ldots, n)
\end{aligned}
$$

LINEAR ALGEBRA

D E T E R M I N A N T

EXAMPLE:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a_{11}} & \boldsymbol{a_{21}} & \boldsymbol{a_{31}} \\ \boldsymbol{a_{12}} & \boldsymbol{a_{22}} & \boldsymbol{a_{32}} \\ \boldsymbol{a_{13}} & \boldsymbol{a_{23}} & \boldsymbol{a_{33}} \end{bmatrix}$$

LINEAR ALGEBRA

D E T E R M I N A N T

ANSWER:

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$$

LINEAR ALGEBRA

I N V E R S E   M A T R I X

The **inverse** of a **square matrix** $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{nxn}}$ is denoted as $\boldsymbol{A}^{-1} \in \mathbb{R}^{\boldsymbol{nxn}}$ and has the following **property**:

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

If $\boldsymbol{A}^{-1}$exists, then it is said that $\boldsymbol{A}$ is **no singular** or **invertible.**

$$(A^{-1})^{-1} = A$$

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

$$(A^{-1})^T = (A^T)^{-1}.$$

LINEAR ALGEBRA

INVERSE MATRIX

The **inverse** of a **square matrix** $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{nxn}}$ can be **calculated** using the **determinant**:

$$\boldsymbol{A^{-1} = \frac{1}{detA} adjA}$$

Where

$$\boldsymbol{adjA = C^T}$$

LINEAR ALGEBRA

EIGEN VALUES AND EIGEN VECTORS

Given a **square matrix** $\boldsymbol{A} \in \mathbb{R}^{nxn}$, it is said that $\boldsymbol{\lambda} \in \mathbb{C}$ is an **eigen value** of $\boldsymbol{A}$ and $\boldsymbol{x} \in \mathbb{C}^{\boldsymbol{n}}$ is an **eigen vector** if:

$$Ax = \lambda x, \qquad x \neq 0$$

$$\downarrow$$

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

LINEAR ALGEBRA

EIGEN VALUES AND EIGEN VECTORS

EXAMPLE:

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{4} & \mathbf{1} \\ \mathbf{-6} & \mathbf{-3} \end{bmatrix}$$

LINEAR ALGEBRA

EIGEN VALUES AND EIGEN VECTORS

ANSWER:

$$\lambda_1 = 3$$
$$\lambda_2 = -2$$

$$x_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

LINEAR ALGEBRA

MATRIX CALCULUS

LINEAR ALGEBRA

THE GRADIENT

Given a **function** $\boldsymbol{f}: \mathbb{R}^{\boldsymbol{mxn}} \rightarrow \mathbb{R}$ that takes as **input** a **matrix** $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{mxn}}$ and **returns** a **real value**.
The **gradient** of $\boldsymbol{f}$ (with respect to $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{mxn}}$) is the **matrix** of **partial derivatives**:

$$\nabla_A f(A) \in \mathbb{R}^{m \times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

Properties:

$$\nabla_x (f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x).$$

$$\text{For } t \in \mathbb{R},\ \nabla_x (t\ f(x)) = t \nabla_x f(x).$$

LINEAR ALGEBRA

T H E   G R A D I E N T

IMPORTANT:

The **gradient** of a **function** is **only defined** if the **function** $\boldsymbol{f}$ is **real-valued**.

Calculating the **gradient** of $\boldsymbol{Ax}, \boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n} \times \boldsymbol{n}}$ with **respect** to the **vector** $\boldsymbol{x}$ is **not defined**, it is **only defined** with respect to **scalars**.

EXAMPLE:

$$f(x,y) = x^2 - xy$$

LINEAR ALGEBRA

THE GRADIENT

ANSWER:

$$\nabla f(x,y) = \begin{bmatrix} 2x - y \\ -x \end{bmatrix}$$

LINEAR ALGEBRA

HESSIAN MATRIX

$$\nabla_x^2 f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

Symmetric property:

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}.$$

LINEAR ALGEBRA
HESSIAN MATRIX

Analogy with **second derivative** of **calculus:**

$$\nabla_x^2 f(x) = \nabla_x (\nabla_x f(x))^T$$

Why the **following** is **wrong?**

$$\nabla_x \nabla_x f(x) = \nabla_x \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

LINEAR ALGEBRA

HESSIAN MATRIX

EXAMPLE:

$$f(x, y) = x^4 y^2$$

LINEAR ALGEBRA

HESSIAN MATRIX

ANSWER:

$$\boldsymbol{f(x,y)} = \begin{bmatrix} \mathbf{12}\boldsymbol{x^2y^2} & \mathbf{8}\boldsymbol{x^3y} \\ \mathbf{8}\boldsymbol{x^3y} & \mathbf{2}\boldsymbol{x^4} \end{bmatrix}$$

LINEAR ALGEBRA

LINEAR AND QUADRATIC FUNCTIONS

EXAMPLE:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{x}$$

Calculate $\nabla_x f(x)$

LINEAR ALGEBRA

LINEAR AND QUADRATIC FUNCTIONS

ANSWER:

$$\boldsymbol{\nabla}_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{b}$$

LINEAR ALGEBRA
QUADRATIC FUNCTIONS

Given a **square matrix** $\boldsymbol{A} \in \mathbb{R}^{\boldsymbol{n\ x\ n}}$ and a vector $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$, the scalar value $\boldsymbol{x^T A x}$ is **denoted** as **quadratic form**.

$$x^T A x = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

LINEAR ALGEBRA

LINEAR AND QUADRATIC FUNCTIONS

Test the following expressions: **(HOMEWORK)**

$$\nabla_x b^T x = b$$

$$\nabla_x x^T A x = 2Ax \text{ (if } A \text{ symmetric)}$$

$$\nabla_x^2 x^T A x = 2A \text{ (if } A \text{ symmetric)}$$