Machine Learning

Week 1: Mathematical Foundations

Jakub Kuwalek

Vrije Universiteit Amsterdam
Faculty of Science

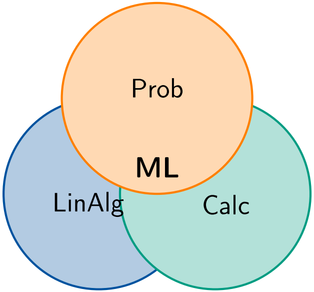
Academic Year 2025/2026

Outline

Why These Preliminaries?

Machine Learning is built on three mathematical pillars:

1. **Linear Algebra** – Vectors, matrices
2. **Calculus** – Derivatives, gradients
3. **Probability Theory** – Expectations



Sigma Notation: The Basics

Definition

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + a_3 + \cdots + a_n$$

Components:

- i – index variable
- 1 – starting value
- n – ending value

Think of it as:

```
for i in range(1,n+1):
    total += a[i]
```

Exercise 1(a): Problem

Given Values

$x_1 = 5, \quad x_2 = 1, \quad x_3 = 4, \quad x_4 = 1, \quad x_5 = 3$

Compute:

$$\sum_i x_i = ?$$

Exercise 1(a): Solution

Answer

$$\sum_i x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

$$= 5 + 1 + 4 + 1 + 3 = \boxed{14}$$

Exercise 1(b): Problem

Given Values

$x_1 = 5, \quad x_2 = 1, \quad x_3 = 4, \quad x_4 = 1, \quad x_5 = 3$

Compute:

$$\sum_i 2 \cdot x_i = ?$$

Exercise 1(b): Solution

Answer

$$\sum_i 2 \cdot x_i = 2(5) + 2(1) + 2(4) + 2(1) + 2(3)$$

$$= 10 + 2 + 8 + 2 + 6 = \boxed{28}$$

Exercise 1(c): Problem

Given Values

$x_1 = 5, \quad x_2 = 1, \quad x_3 = 4, \quad x_4 = 1, \quad x_5 = 3$

Compute:

$$2 \cdot \sum_i x_i = ?$$

Exercise 1(c): Solution

Answer

$$2 \cdot \sum_i x_i = 2 \cdot 14 = \boxed{28}$$

Key Insight

Notice: $\sum_i 2 \cdot x_i = 2 \cdot \sum_i x_i = 28$
This is the **homogeneity property**: constants factor out of sums!

Exercise 1(d): Problem

Given Values

$x_1 = 5, \quad x_2 = 1, \quad x_3 = 4, \quad x_4 = 1, \quad x_5 = 3$

Compute:

$$\sum_{i=1}^{3} x_i = ?$$

Exercise 1(d): Solution

Answer

We only sum the first 3 terms:

$$\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3 = 5 + 1 + 4 = \boxed{10}$$

Exercise 1(e): Problem

Given Values

$x_1 = 5, \quad x_2 = 1, \quad x_3 = 4, \quad x_4 = 1, \quad x_5 = 3$

Compute:

$$\sum_{i=1}^{x_5} x_i = ?$$

Hint: What is the value of x_5?

Exercise 1(e): Solution

Answer

Careful! Since $x_5 = 3$, we sum until index 3:

$$\sum_{i=1}^{x_5} x_i = \sum_{i=1}^{3} x_i = x_1 + x_2 + x_3 = 5 + 1 + 4 = \boxed{10}$$

Common Mistake

The answer is NOT 14. We stop at index 3, not at the element x_5.

Exercise 1(f): Problem

Given Values

$x_1 = 5, \quad x_2 = 1, \quad x_3 = 4, \quad x_4 = 1, \quad x_5 = 3$

Compute:

$$\sum_{i=1}^{3} i \cdot x_i = ?$$

Exercise 1(f): Solution

Answer

Multiply each term by its index:

$$\sum_{i=1}^{3} i \cdot x_i = 1 \cdot x_1 + 2 \cdot x_2 + 3 \cdot x_3$$

$$= 1(5) + 2(1) + 3(4) = 5 + 2 + 12 = \boxed{19}$$

Exercise 2(a): Problem

Question

Is the following **always true**? (c is a constant)

True or False?

$$\sum_i c y_i = c \sum_i y_i$$

Exercise 2(a): Solution

TRUE

$$\sum_i cy_i = cy_1 + cy_2 + \cdots = c(y_1 + y_2 + \cdots) = c\sum_i y_i$$

This is the **homogeneity property** – like factoring out from parentheses.

Exercise 2(b): Problem

Question

Is the following **always true**? (c is a constant)

True or False?

$$\sum_i (c + y_i) = c + \sum_i y_i$$

Exercise 2(b): Solution

FALSE

Left side: c is added **for every term**

Right side: c is added **only once**

Counterexample: Let $n = 2$, $y_1 = 1$, $y_2 = 2$, $c = 10$

Left Side

$(10 + 1) + (10 + 2) = \mathbf{23}$

Right Side

$10 + (1 + 2) = \mathbf{13}$

Exercise 2(c): Problem

Question

Is the following **always true**? (c is a constant)

True or False?

$$\sum_{i=1}^{n}(c + y_i) = nc + \sum_{i=1}^{n} y_i$$

Exercise 2(c): Solution

TRUE

$$\sum_{i=1}^{n}(c + y_i) = \underbrace{c + c + \cdots + c}_{n \text{ times}} + \sum_{i=1}^{n} y_i = nc + \sum_{i=1}^{n} y_i$$

This is the **correct version** of Exercise 2(b)!

Exercise 2(d): Problem

Question

Is the following **always true**?

True or False?

$$\sum_i (x_i + y_i) = \sum_i x_i + \sum_i y_i$$

Exercise 2(d): Solution

TRUE – Additivity Property

$$\sum_i (x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + \cdots$$

Rearranging:

$$= (x_1 + x_2 + \cdots) + (y_1 + y_2 + \cdots) = \sum_i x_i + \sum_i y_i$$

Exercise 2(e): Problem

Question

Is the following **always true**?

True or False?

$$\sum_i \sum_j a_i b_j = \sum_j \sum_i a_i b_j$$

Exercise 2(e): Solution

TRUE

Both compute the sum over all pairs (i, j).

Example: Let $a_1 = 2, a_2 = 3$ and $b_1 = 4, b_2 = 5$

$\sum_i \sum_j a_i b_j$

$= 2(4) + 2(5) + 3(4) + 3(5)$
$= 8 + 10 + 12 + 15 = \mathbf{45}$

$\sum_j \sum_i a_i b_j$

$= 2(4) + 3(4) + 2(5) + 3(5)$
$= 8 + 12 + 10 + 15 = \mathbf{45}$

Exercise 2(f): Problem

Question

Is the following **always true**?

True or False?

$$\sum_i a_i b_i = \left(\sum_i a_i\right)\left(\sum_i b_i\right)$$

Exercise 2(f): Solution

FALSE – This is a common mistake!

Counterexample: $a_1 = 1,\ a_2 = 3,\ b_1 = 4,\ b_2 = 2$

Left Side

$a_1 b_1 + a_2 b_2$
$= 1(4) + 3(2) = \mathbf{10}$

Right Side

$(1+3)(4+2)$
$= 4 \times 6 = \mathbf{24}$

$$10 \neq 24$$

Exercise 2(f): Visual Explanation

Left: $\sum_i a_i b_i$

$a_1 b_1$	$a_1 b_2$
$a_2 b_1$	$a_2 b_2$

Only diagonal!

Right: $(\sum_i a_i)(\sum_j b_j)$

$a_1 b_1$	$a_1 b_2$
$a_2 b_1$	$a_2 b_2$

Entire grid!

What is an Expected Value?

Definition

For outcomes $v_1, v_2, \ldots$ with probabilities $p_1, p_2, \ldots$:

$$\mathbb{E}[V] = \sum_i p_i v_i$$

Example: Fair Die

$$\mathbb{E}[V] = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5$$

The "average" outcome over many rolls.

Exercise 3(a): Problem

Prove the following property:

Homogeneity

$$\mathbb{E}[cV] = c \cdot \mathbb{E}[V]$$

Exercise 3(a): Solution

Proof of Homogeneity

$$
\begin{aligned}
\mathbb{E}[cV] &= \sum_i p_i (c v_i) \\
&= c \sum_i p_i v_i \\
&= c \cdot \mathbb{E}[V] \quad \checkmark
\end{aligned}
$$

Exercise 3(b): Problem

Prove the following property:

Additivity

$$\mathbb{E}[V] + \mathbb{E}[W] = \mathbb{E}[V + W]$$

Exercise 3(b): Solution

Proof of Additivity

$$\begin{aligned}
\mathbb{E}[V] + \mathbb{E}[W] &= \sum_i p_i v_i + \sum_i p_i w_i \\
&= \sum_i (p_i v_i + p_i w_i) \\
&= \sum_i p_i (v_i + w_i) = \mathbb{E}[V + W] \quad \checkmark
\end{aligned}$$

Exercise 3(c): Problem

Show that the following holds:

Statement

$$\mathbb{E}[V] + \mathbb{E}[\sin(V)] = \mathbb{E}[V + \sin(V)]$$

Exercise 3(c): Solution

Answer

This follows directly from the **additivity property**!

Just set $W = \sin(V)$ in the formula:

$$\mathbb{E}[V] + \mathbb{E}[W] = \mathbb{E}[V + W]$$

Therefore:

$$\mathbb{E}[V] + \mathbb{E}[\sin(V)] = \mathbb{E}[V + \sin(V)] \quad \checkmark$$

Exercise 3(d): Problem

What is Variance?

Variance measures how **spread out** values are around the average.

- Low variance: values clustered near the mean
- High variance: values spread far from the mean

Prove the Variance Formula

$$\mathbb{E}[(V - \mathbb{E}[V])^2] = \mathbb{E}[V^2] - (\mathbb{E}[V])^2$$

Exercise 3(d): Understanding the Formula

Left Side: $\mathbb{E}[(V - \mathbb{E}[V])^2]$

- $(V - \mathbb{E}[V])$ = how far each value is from the mean
- $(V - \mathbb{E}[V])^2$ = squared distance (always positive)
- $\mathbb{E}[(V - \mathbb{E}[V])^2]$ = **average squared distance**

Right Side: $\mathbb{E}[V^2] - (\mathbb{E}[V])^2$

- $\mathbb{E}[V^2]$ = expected value of the squared values
- $(\mathbb{E}[V])^2$ = square of the expected value

Variance = "expected square" minus "square of expected"

Exercise 3(d): Solution

Let $\mu = \mathbb{E}[V]$ for convenience.

Proof

$$
\begin{aligned}
\mathbb{E}[(V-\mu)^2] &= \mathbb{E}[V^2 - 2V\mu + \mu^2] \\
&= \mathbb{E}[V^2] - 2\mu\mathbb{E}[V] + \mu^2 \\
&= \mathbb{E}[V^2] - 2\mu^2 + \mu^2 \\
&= \mathbb{E}[V^2] - \mu^2 = \boxed{\mathbb{E}[V^2] - (\mathbb{E}[V])^2}
\end{aligned}
$$

Logarithms: The Basics

Definition

The logarithm is the **inverse of exponentiation**:

$$\text{If } y = b^x, \text{ then } x = \log_b y$$

Examples:

- $\log_{10}(100) = 2$ because $10^2 = 100$
- $\log_{10}(1000) = 3$ because $10^3 = 1000$

Key property: $\log(a \cdot b) = \log a + \log b$

Exercise 4(1): Problem

True or False?

Statement

$$\log_b(b^a) = a$$

Exercise 4(1): Solution

TRUE

If $x = \log_b(b^a)$, then by definition $b^x = b^a$.

Therefore $x = a$. ✓

Intuition: Applying a function then its inverse returns the original.

Exercise 4(2): Problem

True or False?

Statement

$$b^{\log_b a} = a$$

Exercise 4(2): Solution

TRUE

If $x = b^{\log_b a}$, then $\log_b x = \log_b a$.

Therefore $x = a$. ✓

Same reasoning: inverse then function = original.

Exercise 4(3): Problem

True or False?

Statement

$$\log_c a + \log_c b = \log_c (ab)$$

Exercise 4(3): Solution

TRUE – Product Rule

Let $a' = \log_c a$ and $b' = \log_c b$.
Then $c^{a'} = a$ and $c^{b'} = b$.

$$c^{a'} \cdot c^{b'} = c^{a'+b'} \implies a \cdot b = c^{\log_c a + \log_c b}$$

Taking $\log_c$ of both sides: $\log_c(ab) = \log_c a + \log_c b$ ✓

Exercise 4(4): Problem

True or False?

Statement

$$\log(a^b) = b\log(a)$$

Exercise 4(4): Solution

TRUE – Power Rule

$$\log(a^b) = \log(\underbrace{a \cdot a \cdots a}_{b \text{ times}})$$

Using the product rule repeatedly:

$$= \underbrace{\log a + \log a + \cdots}_{b \text{ times}} = b \log a \quad \checkmark$$

Exercise 4(5): Problem

True or False?

Statement

$$\log(a + b) = \log(a) \cdot \log(b)$$

Exercise 4(5): Solution

FALSE – No simple formula for $\log(a + b)$!

Counterexample: Let $a = 10$, $b = 10$

Left Side

$\log_{10}(10 + 10)$
$= \log_{10}(20) \approx \mathbf{1.3}$

Right Side

$\log_{10}(10) \cdot \log_{10}(10)$
$= 1 \cdot 1 = \mathbf{1}$

$$1.3 \neq 1$$

Exercise 4(6): Problem

True or False?

Statement

$$\log(a/b) = \log a - \log b$$

Exercise 4(6): Solution

TRUE – Quotient Rule

$$\begin{aligned}\log(a/b) &= \log(a \cdot b^{-1}) \\ &= \log a + \log(b^{-1}) \\ &= \log a + (-1)\log b \\ &= \log a - \log b \quad \checkmark\end{aligned}$$

Vectors and Matrices: Notation

Vector $\mathbf{x} \in \mathbb{R}^n$

A list of n numbers:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$

An $m \times n$ grid:

$$\mathbf{W} = \begin{pmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{m1} & \cdots & W_{mn} \end{pmatrix}$$

Exercise 5(a): Problem

Explain in words:

Notation

$$f : \mathbb{R}^3 \to \mathbb{R}^2$$

Exercise 5(a): Solution

Answer

f is a **function** that:

- Takes a **3-dimensional vector** as input
- Outputs a **2-dimensional vector**

Example: A camera projecting a 3D scene onto a 2D image.

Exercise 5(a): Visual Example

3D Vector (Input)

$$x = \begin{pmatrix} 1.5 \\ 2.0 \\ 3.7 \end{pmatrix} \in \mathbb{R}^3$$

$$\xrightarrow{\quad f \quad}$$

2D Vector (Output)

$$y = \begin{pmatrix} 4.2 \\ 1.8 \end{pmatrix} \in \mathbb{R}^2$$

Interpretation

The function f transforms a point in 3D space to a point in 2D space.

Exercise 5(b): Problem

Explain in words:

Notation

$$y = Wx$$

Exercise 5(b): Solution

Answer

The vector y is the result of **matrix-vector multiplication**.

Matrix W transforms vector x into vector y.

This is a linear transformation of x.

Exercise 5(b): Visual Example

Matrix W		**Vector** x		**Vector** y
$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$	$\times$	$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$	$=$	$\begin{pmatrix} 14 \\ 32 \end{pmatrix}$
2×3		3×1		2×1

How it works

$$y_1 = 1(1) + 2(2) + 3(3) = 14$$
$$y_2 = 4(1) + 5(2) + 6(3) = 32$$

Exercise 5(c): Problem

Explain in words:

Notation

$$z = \mathrm{y}^\top \mathrm{x}$$

Exercise 5(c): Solution

Answer

z is the **dot product** of vectors y and x.

$$z = \sum_i y_i x_i = y_1 x_1 + y_2 x_2 + \cdots$$

The result is a **scalar** (single number), not a vector!

Exercise 5(c): Example with Numbers

$$\underbrace{\begin{pmatrix} 2 & 3 & 4 \end{pmatrix}}_{y^\top} \times \begin{pmatrix} 5 \\ 1 \\ 2 \end{pmatrix} = \underbrace{21}_{z}$$

Calculation

$$z = y^\top x = 2(5) + 3(1) + 4(2) = 10 + 3 + 8 = \boxed{21}$$

Multiply corresponding elements, then sum them all!

Exercise 5(d): Problem

Explain in words:

Notation

$$W \in \mathbb{R}^{5\times 4}$$

Exercise 5(d): Solution

Answer

W is a matrix with:

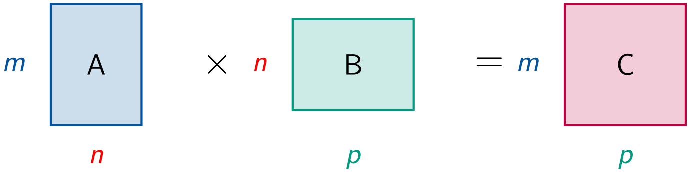
- **5 rows**
- **4 columns**
- Real-valued elements

Matrix Multiplication: The Rule

Dimension Rule

$$(m \times n) \cdot (n \times p) = (m \times p)$$

Inner dimensions must match!



Matrix Multiplication: Examples with Numbers

Valid: $(2 \times 3) \cdot (3 \times 2)$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \times \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 22 & 28 \\ 49 & 64 \end{pmatrix}$$

Inner dimensions match: $3 = 3$ ✓

Invalid: $(2 \times 3) \cdot (2 \times 2)$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \times \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \text{impossible!}$$

Inner dimensions do NOT match: $3 \neq 2$

Exercise 6(1): Problem

Is this operation possible?

Matrix Multiplication

Multiply a (5×4) matrix by another (5×4) matrix

Exercise 6(1): Solution

IMPOSSIBLE

$$(5 \times 4) \cdot (5 \times 4)$$

Inner dimensions: $4 \neq 5$

The number of columns in the first matrix must equal the number of rows in the second.

Exercise 6(2): Problem

Is this operation possible?

Element-wise Multiplication

Element-wise multiply a (5×4) matrix by another (5×4) matrix

Exercise 6(2): Solution

POSSIBLE

For element-wise operations, dimensions must be **identical**.

$(5 \times 4) \odot (5 \times 4)$ – Same dimensions! $\checkmark$

Example with Numbers

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \odot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 5 & 12 \\ 21 & 32 \end{pmatrix}$$

Each element multiplied with the element at the same position.

Exercise 6(3): Problem

Is this operation possible?

Matrix Multiplication

Multiply a (5×4) matrix by a (4×5) matrix

Exercise 6(3): Solution

POSSIBLE

$$(5 \times 4) \cdot (4 \times 5) = (5 \times 5)$$

Inner dimensions match: $4 = 4$ ✓

Result is a 5×5 matrix.

Exercise 6(6): Problem

Is this operation possible?

Statement

Multiply **any** matrix by its transpose

Exercise 6(6): Solution

ALWAYS POSSIBLE

If A is $(m \times n)$, then $A^\top$ is $(n \times m)$.

$$(m \times n) \cdot (n \times m) = (m \times m)$$

Inner dimensions always match! Result is always square.

Exercise 8(a): Problem

What is a Dot Product?

The **dot product** combines two vectors into a single number:

1. Multiply corresponding elements together
2. Add all the products

$$\mathbf{w}^\top \mathbf{x} = w_1 x_1 + w_2 x_2 + w_3 x_3 + \cdots$$

Write as a dot product:

$$f(a, b, c) = \alpha a + \beta b + \gamma c$$

Exercise 8(a): Solution

Answer

Let $\mathrm{x} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and $\mathrm{w} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$

Then:

$$f(\mathrm{x}) = \mathrm{w}^\top \mathrm{x} = \alpha a + \beta b + \gamma c \quad \checkmark$$

Exercise 8(b): Problem

Write as a dot product:

Affine Function (with constant)

$$f(a, b, c) = \alpha a + \beta b + \gamma c + \delta$$

Exercise 8(b): Solution

The Bias Trick – Add a 1!

Let $\mathrm{x} = \begin{pmatrix} a \\ b \\ c \\ 1 \end{pmatrix}$ and $\mathrm{w} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}$

Then:

$$f(\mathrm{x}) = \mathrm{w}^\top \mathrm{x} = \alpha a + \beta b + \gamma c + \delta \cdot 1 \quad \checkmark$$

Exercise 8(c): Problem

Write as $x^\top W x$:

Quadratic Function

$$f(a,b) = \alpha a^2 + \beta ab + \gamma ba + \delta b^2$$

Exercise 8(c): Solution

Quadratic Form

Let $\mathrm{x} = \begin{pmatrix} a \\ b \end{pmatrix}$ and $\mathrm{W} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$

$$
\begin{aligned}
\mathrm{x}^\top \mathrm{W} \mathrm{x} &= \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} \alpha a + \beta b \\ \gamma a + \delta b \end{pmatrix} \\
&= \alpha a^2 + \beta ab + \gamma ba + \delta b^2 \quad \checkmark
\end{aligned}
$$

Derivatives: Notation

Two Common Notations

Lagrange: $f'(x)$

Leibniz: $\dfrac{df}{dx}$

Chain Rule (why Leibniz is useful):

$$\frac{d}{dx}[g(f(x))] = \frac{dg}{df} \cdot \frac{df}{dx}$$

Exercise 9(a): Problem

Given:

$$f(x) = 3x^2 + 5x + 1$$

Find:

$$\frac{df}{dx} = ?$$

Exercise 9(a): Solution

Using the power rule: $\frac{d}{dx}x^n = nx^{n-1}$

$$\frac{df}{dx} = \frac{d}{dx}(3x^2) + \frac{d}{dx}(5x) + \frac{d}{dx}(1)$$
$$= 6x + 5 + 0 = \boxed{6x + 5}$$

Exercise 9(b): Problem

Given:

$$f(x) = 3x^2 + 5x + 1$$

Find:

For which x is $f(x)$ at its **minimum**?

Exercise 9(b): Solution

Why does $\frac{df}{dx} = 0$ at minimum?

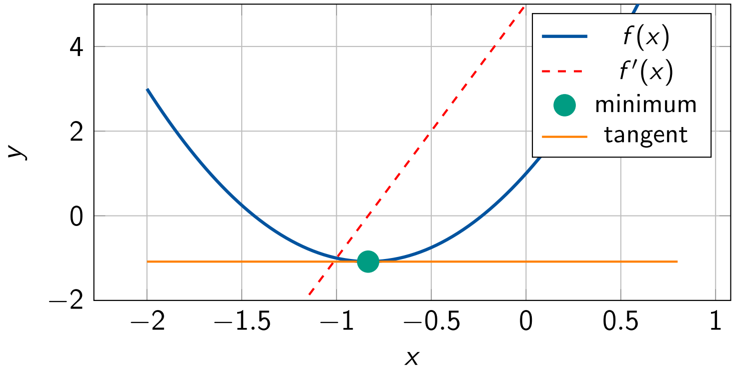
The derivative tells us the **slope** of the function at each point.

- Slope > 0: function is going UP
- Slope < 0: function is going DOWN
- Slope $= 0$: function is **flat** (minimum or maximum!)

Calculation

$$\frac{df}{dx} = 0 \implies 6x + 5 = 0 \implies x = \boxed{-\frac{5}{6}}$$

Exercise 9(b): Visual Explanation



At minimum: slope $= 0$ (flat tangent line)

Exercise 9(c): Problem

Given:

$$f(x) = 3x^2 + 5x + 1 \quad \text{and} \quad \frac{dg}{dx} = \frac{\sin x}{x}$$

Find:

Let $h(x) = g(f(x))$. What is $\frac{dh}{dx}$?

Exercise 9(c): Solution

Apply the Chain Rule

$$\frac{dh}{dx} = \frac{dg(f(x))}{df(x)} \cdot \frac{df(x)}{dx}$$

We know: $\frac{dg}{df} = \frac{\sin(f(x))}{f(x)}$ and $\frac{df}{dx} = 6x + 5$

$$\frac{dh}{dx} = \boxed{\frac{\sin(3x^2 + 5x + 1)}{3x^2 + 5x + 1} \cdot (6x + 5)}$$

Exercise 10(a): Problem

Given:

$f(\mathrm{x}) = 3x_1^2 + 4x_1x_2 - x_2^2 \quad \text{where } \mathrm{x} \in \mathbb{R}^2$

Find:

$\frac{\partial f}{\partial x_1} = ?$

Exercise 10(a): Solution

Treat x_2 as a constant

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1}(3x_1^2) + \frac{\partial}{\partial x_1}(4x_1x_2) - \frac{\partial}{\partial x_1}(x_2^2)$$
$$= 6x_1 + 4x_2 - 0 = \boxed{6x_1 + 4x_2}$$

Exercise 10(b): Problem

Given:

$f(\mathrm{x}) = 3x_1^2 + 4x_1x_2 - x_2^2 \quad \text{where } \mathrm{x} \in \mathbb{R}^2$

Find:

$\frac{\partial f}{\partial x_2} = ?$

Exercise 10(b): Solution

Treat x_1 as a constant

$$\frac{\partial f}{\partial x_2} = \frac{\partial}{\partial x_2}(3x_1^2) + \frac{\partial}{\partial x_2}(4x_1 x_2) - \frac{\partial}{\partial x_2}(x_2^2)$$
$$= 0 + 4x_1 - 2x_2 = \boxed{4x_1 - 2x_2}$$

Exercise 10(c): Problem

Given:

$f(x) = 3x_1^2 + 4x_1x_2 - x_2^2 \quad \text{where } x \in \mathbb{R}^2$

Find:

What is the gradient $\nabla f(x)$?

Exercise 10(c): Solution

The Gradient

The gradient is the vector of all partial derivatives:

$$\nabla f(\mathrm{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \boxed{\begin{pmatrix} 6x_1 + 4x_2 \\ 4x_1 - 2x_2 \end{pmatrix}}$$

Exercise 10(d): Problem

Given:

$\nabla f : \mathbb{R}^2 \to ?$

Find:

What are the **domain** and **range** of ∇f?

Exercise 10(d): Solution

Domain and Range

Domain: $\mathbb{R}^2$ (same as input to f)

Range: $\mathbb{R}^2$ (gradient has same dimension as input)

$\nabla f : \mathbb{R}^2 \to \mathbb{R}^2$

Key Insight: Gradient Descent

The Gradient Points Uphill

∇f points in the direction of **steepest increase**.

In Machine Learning

To **minimize** a loss function, go **opposite** to the gradient: $\mathsf{x}_{\text{new}} = \mathsf{x}_{\text{old}} - \eta \nabla f$
where η is the learning rate.

Thank you!

Questions?

Jakub Kuwalek
Vrije Universiteit Amsterdam