Artificial Intelligence I 2024/2025

Week 2 Tutorial and Additional Exercises

Differentiation, Partial Derivatives and Gradients

School of Computer Science

20th January 2025

In this tutorial...

In this tutorial we will be covering

- Functions and compositions.
- Linear and polynomial functions and their vector notations
- Sigmoid function
- Derivatives
- Partial derivatives
- Chain rule
- Gradient vectors.
- Exercises on all the above.

Functions

We revisit functions. Recall the definition:

Definition 1 (Function)

Let A and B be sets of real numbers. A *function* $f : A \to B$ is a rule that maps each element in A to exactly one element in B. The unique element that x is mapped into, by the function f, is called the *image* of x under f, and is denoted $f(x)$.

- We will focus on functions that have a closed-form expression.
- Some examples are

$$f(x) = x^3 - 1, \qquad g(x) = \sin(2x), \qquad h(x) = e^{-x^2}.$$

Compositions of functions

Definition 2 (Function composition)

If $f : A \to B$ and $g : B \to C$, then their *composition*, $g \circ f : A \to C$ (read "g of f"), is defined as

$$(g \circ f)(x) = g(f(x))$$

wherever it exists.

- Schematically:

$$\begin{tikzcd} & f(x) \arrow[dr, "g"] & \\ x \arrow[ur, "f"] \arrow[rr, bend right, "g \circ f"'] & & g(f(x)) \end{tikzcd}$$

- Compositions can be extended to three or more functions.

Exercise 1

In each case, use definition 2 to find $(g \circ f)(x)$ and $(f \circ g)(x)$:

1. $f(x) = x^2$ and $g(x) = \cos(x)$.
2. $f(x) = e^x$ and $g(x) = x^3$.
3. $f(x) = -3x$ and $g(x) = \ln(x)$.
4. $f(x) = \sin(x)$ and $g(x) = \frac{1}{x}$.

Exercise 2

In each case, use definition 2 to write f as a composition of two or more elementary functions (see middle column in table 1 for the elementary functions):

1. $f(x) = 2^{x^3}$.
2. $f(x) = \frac{1}{\sin(x)}$.
3. $f(x) = \sqrt{\ln(3x)}$.
4. $f(x) = e^{-3\sin(x^2)}$.

Some examples of functions

1. A *linear function* has all variables raised to the power of 1.
   - Univariate linear function: $f(x) = a_0 + a_1 x$, where $a_0, a_1 \in \mathbb{R}$.
   - Multivariate linear function: $f(x_1, x_2, \cdots, x_n) = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$, where $a_0, a_1, \ldots, a_n \in \mathbb{R}$.
   - Multivariate linear function (summation notation): $f(x_1, x_2, \ldots, x_n) = \sum_{i=0}^{n} a_i x_i$.
   - Multivariate linear function (vector notation): $f(\mathbf{x}) = a_0 + \mathbf{a}^T \mathbf{x}$, where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$ and $\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T$.
2. A *polynomial function* of degree n has only integer powers (from 0 to n) of its variables.
   - Univariate polynomial function: $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$, where $a_0, a_1, \ldots, a_n \in \mathbb{R}$.
   - Univariate polynomial function (summation notation): $f(x) = \sum_{i=0}^{n} a_i x^i$.

Exercise 3

In each of the following examples, determine whether the given function is linear in $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$:

1. $f(\mathbf{x}) = 2 + 3x_1 + 5x_2$;
2. $f(\mathbf{x}) = 7 - \sqrt{2}x_1$;
3. $f(\mathbf{x}) = x_1^2 + x_2$;
4. $f(\mathbf{x}) = 5$.

Exercise 4

In each of the following examples, determine whether the given function is polynomial in x, and the degree of the polynomial:

1. $f(x) = 2 + x - x^2$;
2. $f(x) = 2 + \sqrt{x}$;
3. $f(x) = -2x + x^3$;
4. $f(x) = -7$.

Sigmoid function

1. The *sigmoid function* $\sigma : \mathbb{R} \to (0, 1)$ is defined as
$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$
2. $\sigma(x)$ is strictly increasing in x (prove it!).
3. $\lim_{x \to -\infty} \sigma(x) = 0$ and $\lim_{x \to \infty} \sigma(x) = 1$ (prove it!).
4. The sigmoid function has the useful property of mapping any real number to the interval $(0, 1)$ (prove it!).
5. Therefore, it finds applications to classification problems, as we will see later in the course.

Exercise 5

Prove the following properties of the sigmoid function $\sigma(x)$:

1. $\sigma(x) = 1 - \sigma(-x)$;
2. $\sigma'(x) = \sigma(x)(1 - \sigma(x))$.
3. $\sigma^{-1}(y) = \ln(\frac{y}{1-y})$.

Derivatives

Definition 3 (Derivative)

The *derivative* of a function f is another function f' defined as

$$f'(x) = \lim_{t \to 0} \frac{f(x+t) - f(x)}{t}$$

wherever the limit exists.

- The derivative f' is also written as $\frac{df}{dx}$.
- For a point x^*, the quantity $f'(x^*)$ is the instantaneous rate of change of f at x^*.
- The limit in definition 3 is of the form $\frac{0}{0}$. To calculate $f'(x)$ we expand the formulas in the numerator to eliminate the denominator, before applying the limit.

Derivatives (continued)

Geometrically, $f'(x^*)$ equals the slope of a line that is tangent to the graph of f at point $(x^*, f(x^*))$, as shown in figure 1.

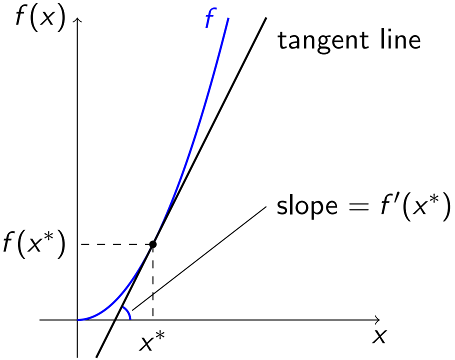


Figure 1: Slope of a function f at a point x^*.

Derivative rules I

Name	$f(x)$	$f'(x)$
constant	c	0
linear	cx	c
power	$x^c,\ c \neq 0$	cx^{c-1}
exponential	$c^x,\ c > 0$	$c^x \ln(c)$
logarithmic	$\log_c(x),\ 0 < c \neq 1$	$\frac{1}{x \ln(c)}$
sine	$\sin(x)$	$\cos(x)$
cosine	$\cos(x)$	$-\sin(x)$

Table 1: Derivatives of elementary functions (c is a constant).

We next prove some of these rules using definition 3. Try to prove the others yourself!

Derivations of table 1

- Let $f(x) = c$, where c is a constant. Then

$$f'(x) = \lim_{t \to 0} \frac{f(x+t) - f(x)}{t} = \lim_{t \to 0} \frac{c - c}{t} = \lim_{t \to 0} 0 = 0.$$

- Let $f(x) = cx$, where c is a constant. Then

$$f'(x) = \lim_{t \to 0} \frac{f(x+t) - f(x)}{t} = \lim_{t \to 0} \frac{c(x+t) - cx}{t} = \lim_{t \to 0} c = c.$$

- Let $f(x) = x^c$, where $c = 2$. Then

$$f'(x) = \lim_{t \to 0} \frac{f(x+t) - f(x)}{t} = \lim_{t \to 0} \frac{(x+t)^2 - x^2}{t} =$$

$$\lim_{t \to 0} \frac{x^2 + 2xt + t^2 - x^2}{t} = \lim_{t \to 0} \frac{2xt + t^2}{t} = \lim_{t \to 0} (2x+t) = 2x.$$

Exercise 6

In each case, use table 1 to find $f'(x)$:

1. $f(x) = 5$.
2. $f(x) = -3x$.
3. $f(x) = x^4$.
4. $f(x) = \sqrt{x}$.
5. $f(x) = \frac{1}{x}$.
6. $f(x) = e^x$.
7. $f(x) = \ln(x)$.

Derivative rules II

Rule	Function	Derivative
Constant multiple	$c \cdot f$	$c \cdot f'$
Sum	$f + g$	$f' + g'$
Product	$f \cdot g$	$f' \cdot g + f \cdot g'$
Quotient	$\frac{f}{g}$	$\frac{f' \cdot g - f \cdot g'}{g^2}$
Composition	$f \circ g$	$(f' \circ g) \cdot g'$

Table 2: Derivative rules for functions f and g, provided all shown derivatives exist (c is a constant).

Exercise 7

In each case, use table 2 to find $f'(x)$:

1. $f(x) = -4x^3$.
2. $f(x) = x^3 + x^2$.
3. $f(x) = x \cdot e^x$.
4. $f(x) = \frac{\sin(x)}{x}$.
5. $f(x) = e^{\sin(x)}$.
6. $f(x) = \ln(x^3 + x^2)$.
7. $f(x) = \cos(x^3 - 3x^2)$.
8. $f(x) = e^{-4x^2+5x+8}$.

Multivariate functions in general

Functions of several variables are defined analogously to functions in the single-variable case.

Definition 4 (Multivariate function)

Let A be a set of n-tuples of real numbers and B be a set of real numbers. A *function of n-variables* $f : A \to B$ is a rule that maps each element in A to exactly one element in B. The unique element that $(x_1, x_2, \ldots, x_n)$ is mapped into, by the function f, is called the *image* of $(x_1, x_2, \ldots, x_n)$ under f, and is denoted $f(x_1, x_2, \ldots, x_n)$.

- Some examples are

$$f(x_1, x_2) = x_1 + 2x_2, \quad g(x_1, x_2, x_3) = x_1 e^{x_2} + \ln(x_3).$$

- We usually group all variables into a vector and write $f(\mathbf{x})$.

Partial derivatives

When we hold all but one variable of a function constant and take the derivative with respect to that one variable, we get a partial derivative.

Definition 5 (Partial derivative)

The *partial derivative* of a function f of n-variables $(x_1, \ldots, x_n)$ with respect to x_i $(1 \leq i \leq n)$ is another function $\frac{\partial f}{\partial x_i}$ defined as

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) = \lim_{t \to 0} \frac{f(x_1, \ldots, x_i + t, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{t}$$

wherever the limit exists.

The quantity $\frac{\partial f}{\partial x_i}(x_1^*, \ldots, x_n^*)$ is the instantaneous rate of change of f at point $(x_1^*, \ldots, x_n^*)$ when moving parallel to the i-th axis.

Exercise 8

In each case, find $\frac{\partial f}{\partial x_1}(x_1, x_2)$ and $\frac{\partial f}{\partial x_2}(x_1, x_2)$.

1. $f(x_1, x_2) = 2x_1^2 - 3x_2 - 4$.
2. $f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2$.
3. $f(x_1, x_2) = (x_1 x_2 - 1)^2$.
4. $f(x_1, x_2) = (2x_1 - 3x_2)^2$.
5. $f(x_1, x_2) = e^{x_1 x_2 + 1}$.
6. $f(x_1, x_2) = \ln(x_1 + x_2)$.

Chain rule

Theorem 6 (Chain rule)

If f is a function of g and g is a function of : $y = f(g(x))$ $(u = g(x))$, then

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = f'(g(x))g'(x).$$

Proof.

The proof is out of the scope of this module! If interested, please refer to an elementary calculus textbook! $\square$

Exercise 9

In each case write $\frac{df}{dt}$ as a function of t, using theorem 6.

1. $f(x_1, x_2) = x_1^2 - x_2^2$, where $x_1 = t^2$, $x_2 = -2t^3$.
2. $f(x_1, x_2) = x_1^2 + x_2^2$, where $x_1 = \cos(t)$, $x_2 = \sin(t)$.
3. $f(x_1, x_2) = 2x_2 e^{x_1}$, where $x_1 = \ln(t^2 + 1)$, $x_2 = t^2$.
4. $f(x_1, x_2) = \sin(x_1 x_2)$, where $x_1 = t$, $x_2 = \ln(t)$.

Gradient vectors

We now define the vector of partial derivatives of a multivariate function f.

Definition 7 (Gradient vector)

The *gradient vector* of a function f of n-variables $(x_1, \ldots, x_n)$ is the vector-valued function ∇f, read "del f", defined as

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T .$$

∇f is a vector-valued function. When evaluated at point $(x_1^*, \ldots, x_n^*)$, its direction shows the *direction of greatest increase* of f from point $(x_1^*, \ldots, x_n^*)$, and its norm equals the directional derivative of f along that direction.

Exercise 10

Given a univariate linear function $f(x) = w_0 + w_1 x$, we define another function: $g(w_0, w_1) = (f - y)^2$, what is the gradient vector of g with respect to $\mathbf{w} = \begin{bmatrix} w_0 & w_1 \end{bmatrix}^T$?

Hint: In function g, we treat w_0 and w_1 as independent variables, and thus x and y are constant.

Exercise 11

Let $\mathbf{w} = \begin{bmatrix} w_0 & w_1 \end{bmatrix}^T$. Given the function $h(\mathbf{w}) = \frac{1}{1+e^{-(w_0+w_1x)}}$, we define:

1. $f(\mathbf{w}) = -\ln(h(\mathbf{w}))$
2. $g(\mathbf{w}) = -\ln(1 - h(\mathbf{w}))$

What are the gradient vectors $\nabla f(\mathbf{w})$ and $\nabla g(\mathbf{w})$ with respect to the vector $\mathbf{w}$?

Some applications of gradient vectors

Gradient vectors find applications in various fields. Here is a few:

- In minimisation problems, if f is the objective function and $\mathbf{x}$ is the current solution, we usually follow the update rule

  $$\mathbf{x} \leftarrow \mathbf{x} - \alpha \nabla f(\mathbf{x})$$

  where $\alpha > 0$. This is because the objective function at $\mathbf{x}$ decreases the most along the direction of $-\nabla f(\mathbf{x})$.
- When $f(\mathbf{x})$ denotes the potential energy at point $\mathbf{x}$, the direction of $-\nabla f(\mathbf{x})$ shows the flow of particles, as this direction reduces their potential energy the quickest. This applies to electrostatics, fluid flow, gravitation and heat flow problems and shows the direction of particles or objects.
- Gradient vectors also possess nice geometric properties.

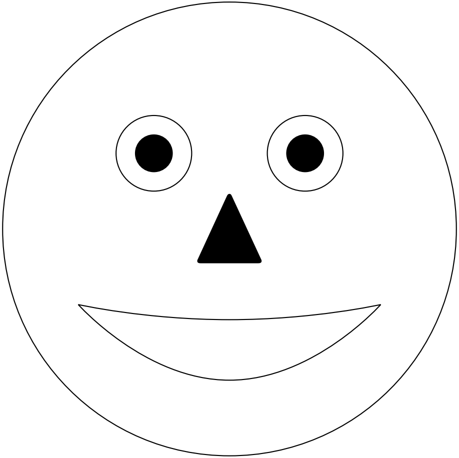
Resources

Some extra resources about functions and derivatives:

- Interactive derivative plotter: `https://www.mathsisfun.com/calculus/derivatives-introduction.html`
- Graphical tool: `https://geogebra.org/calculator`

Any questions?

Until the next time...



Thank you for your attention!