Artificial Intelligence I 2023/2024

Week 2 Tutorial and Additional Exercises

Differentiation, Partial Derivatives and Gradients

School of Computer Science

February 13, 2024

In this tutorial...

In this tutorial we will be covering

- Functions and compositions.
- Derivatives
- Partial derivatives
- Chain rule
- Directional derivatives.
- 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 1: Solution

There are more than one possible solutions. Here is one (in the same order):

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

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. $f(x) = e^{-3\sin(x^2)}$.

Exercise 2: Solution

There are more than one possible solutions. Here is one (in the same order):

1. $f = f_2 \circ f_1$ where $f_1(x) = x^3$, $f_2(x) = 2^x$.
2. $f = f_2 \circ f_1$ where $f_1(x) = \sin(x)$ and $f_2(x) = \frac{1}{x}$.
3. $f = f_3 \circ f_2 \circ f_1$ where $f_1(x) = 3x$, $f_2(x) = \ln(x)$, $f_3(x) = \sqrt{x}$.
4. $f = f_4 \circ f_3 \circ f_2 \circ f_1$ where $f_1(x) = x^2$, $f_2(x) = \sin(x)$, $f_3(x) = -3x$, $f_4(x) = e^x$.

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.

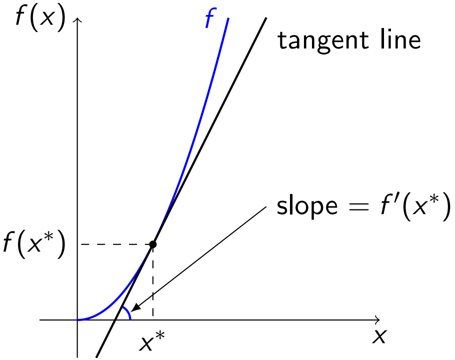


Figure 1: Slope of a function f at a point x^*.

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!

(OPTIONAL) 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 3

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. $f(x) = \ln(x)$.

Exercise 3: Solution

The derivatives are (in the same order):

1. $f'(x) = 0$.
2. $f'(x) = -3$.
3. $f'(x) = 4x^3$.
4. $f'(x) = \frac{1}{2\sqrt{x}}$.
5. $f'(x) = -\frac{1}{x^2}$.
6. $f'(x) = e^x$.
7. $f'(x) = \frac{1}{x}$.

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).

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

(OPTIONAL) Derivations of table 2

- Let f be a function and c be a constant. Then

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

- Let f and g be functions. Then

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

Exercise 4

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. $f(x) = e^{-4x^2+5x+8}$.

Exercise 4: Solution

The derivatives are (in the same order):

1. $f'(x) = -12x^2$.
2. $f'(x) = 3x^2 + 2x$.
3. $f'(x) = e^x + x \cdot e^x$.
4. $f'(x) = \frac{\cos(x) \cdot x - \sin(x)}{x^2}$.
5. $f'(x) = \cos(x) \cdot e^{\sin(x)}$.
6. $f'(x) = \frac{3x+2}{x^2+x}$.
7. $f'(x) = -\sin(x^3 - 3x^2) \cdot (3x^2 - 6x)$.
8. $f'(x) = e^{-4x^2+5x+8} \cdot (-8x + 5)$.

Functions of several variables

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

Definition 4 (Multi-variable 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

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

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 5

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. $f(x_1, x_2) = \ln(x_1 + x_2)$.

Exercise 5: Solution

The partial derivatives are (in the same order):

1. $\frac{\partial f}{\partial x_1}(x_1, x_2) = 4x_1$ and $\frac{\partial f}{\partial x_2}(x_1, x_2) = -3$.
2. $\frac{\partial f}{\partial x_1}(x_1, x_2) = 2x_1 - x_2$ and $\frac{\partial f}{\partial x_2}(x_1, x_2) = -x_1 + 2x_2$.
3. $\frac{\partial f}{\partial x_1}(x_1, x_2) = 2x_1x_2^2 - 2x_2$ and $\frac{\partial f}{\partial x_2}(x_1, x_2) = 2x_1^2x_2 - 2x_1$.
4. $\frac{\partial f}{\partial x_1}(x_1, x_2) = 8x_1 - 12x_2$ and $\frac{\partial f}{\partial x_2}(x_1, x_2) = 18x_2 - 12x_1$.
5. $\frac{\partial f}{\partial x_1}(x_1, x_2) = x_2 \cdot e^{x_1x_2+1}$ and $\frac{\partial f}{\partial x_2}(x_1, x_2) = x_1 \cdot e^{x_1x_2+1}$.
6. $\frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{1}{x_1+x_2}$ and $\frac{\partial f}{\partial x_2}(x_1, x_2) = \frac{1}{x_1+x_2}$.

Chain rule

- Suppose we have a function f of n variables $(x_1, \ldots, x_n)$ and each of them is itself a function of the same variable t, that is

$$x_1 = x_1(t), \quad x_2 = x_2(t), \quad \ldots \quad x_n = x_n(t).$$

- We can think of the input point $(x_1, \ldots, x_n)$ of f tracing a curve in space, which is parameterised by t.
- Then, f can be seen as a single-variable function of t, as all of its variables are themselves functions of t.
- The *chain rule* allows us to find the derivative of f with respect to t, that is $\frac{df}{dt}$. It is a generalization of the composition rule in table 2.
- We next present the formal theorem of the chain rule.

Chain rule (continued)

Theorem 6 (Chain rule)

If f is a function of n-variables $(x_1, x_2 \ldots, x_n)$ and if $x_1 = x_1(t), x_2 = x_2(t), \ldots, x_n = x_n(t)$ are functions of t, then

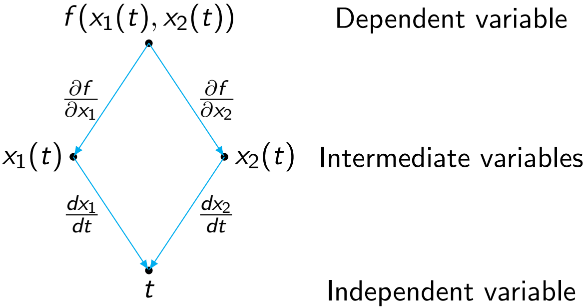
$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}.$$

Proof.

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

Chain rule (continued)

Figure 2 is a mnemonic for the chain rule of 2-variable functions.
Try to draw it for 3 or more variable functions!



$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{d_t} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt}$$

Figure 2: To find $\frac{df}{dt}$, start at f and read down each route to t, multiplying derivatives along the way. Then add the products.

Exercise 6

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 $x_1 = t$, $x_2 = \ln(t)$.

Exercise 6: Solution

The derivatives are (in the same order):

1. $\frac{df}{dt} = 2x_1(2t) + (-2x_2)(-6t^2) = 4t^3 - 24t^5$.
2. $\frac{df}{dt} = 2x_1 \cdot (-\sin(t)) + 2x_2 \cdot \cos(t) =$
$-2\cos(t)\sin(t) + 2\sin(t)\cos(t) = 0$.
3. $\frac{df}{dt} = 2x_2 e^{x_1} \frac{2t}{t^2+1} + 2e^{x_1}(2t) =$
$2t^2(t^2+1)\frac{2t}{t^2+1} + 2(t^2+1)2t = 8t^3 + 4t$.
4. $\frac{df}{dt} = x_2 \cdot \cos(x_1 x_2) + x_1 \cdot \cos(x_1 x_2)\frac{1}{t} =$
$\ln(t)\cos(t \cdot \ln(t)) + \cos(t \cdot \ln(t))$.

Directional derivatives

Directional derivatives of multi-variable functions are generalizations of partial derivatives along arbitrary directions.

Definition 7 (Directional derivative)

The *directional derivative* of a function f of n-variables $(x_1, \ldots, x_n)$ in the direction of the unit vector $\mathbf{u} = (u_1, \ldots, u_n)$ is another function $D_\mathbf{u} f$ defined as

$$D_\mathbf{u} f(x_1, \ldots, x_n) = \lim_{t \to 0} \frac{f(x_1 + tu_1, \ldots, x_n + tu_n) - f(x_1, \ldots, x_n)}{t}$$

wherever the limit exists.

The quantity $D_\mathbf{u} f(x_1^*, \ldots, x_n^*)$ is the instantaneous rate of change of f at point $(x_1^*, \ldots, x_n^*)$ when moving along the direction of $\mathbf{u}$.

Gradient vectors

To calculate $D_{\mathbf{u}} f(x_1, \ldots, x_n)$, we write each variable of f as a function of t:

$$x_1(t) = x_1 + tu_1, \quad x_2(t) = x_2 + tu_2, \ldots \quad x_n(t) = x_n + tu_n$$

and calculate $\frac{df}{dt}$ using theorem 6. We obtain:

$$\begin{aligned}
D_{\mathbf{u}} f(x_1, \ldots, x_n) = \frac{df}{dt} &= \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt} \\
&= \frac{\partial f}{\partial x_1} u_1 + \frac{\partial f}{\partial x_2} u_2 + \cdots + \frac{\partial f}{\partial x_n} u_n \\
&= \underbrace{\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)}_{\nabla f} \cdot \underbrace{(u_1, u_2, \ldots, u_n)^T}_{\mathbf{u}}.
\end{aligned}$$

Gradient vectors (continued)

We now define the vector of partial derivatives of f.

Definition 8 (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

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right).$$

Theorem 9 (Directional derivative as inner product)

Let f be a multi-variable function and $\mathbf{u}$ be a unit vector. Then

$$D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u}.$$

Gradient vectors (continued)

Some further notes about ∇f.

- Since $\mathbf{u}$ is a unit vector, we have

$$D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u} = \|\nabla f\| \, \|\mathbf{u}\| \cos(\theta_{\nabla f, \mathbf{u}}) = \|\nabla f\| \cos(\theta_{\nabla f, \mathbf{u}})$$

  where $\theta_{\nabla f, \mathbf{u}}$ is the angle between ∇f and $\mathbf{u}$. Therefore, $D_{\mathbf{u}} f$ is the largest if $\mathbf{u}$ points towards the direction of ∇f, the smallest if $\mathbf{u}$ points towards the direction of $-\nabla f$, and equals 0 if $\mathbf{u}$ and ∇f are perpendicular.

- ∇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 7

In each case, find $D_{\mathbf{u}} f(x_1, x_2)$ using theorem 9.

1. $f(x_1, x_2) = 2x_1^2 - 3x_2 - 4$, $\mathbf{u} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.
2. $f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2$, $\mathbf{u} = (-\frac{\sqrt{3}}{2}, \frac{1}{2})$.
3. $f(x_1, x_2) = (x_1 x_2 - 1)^2$, $\mathbf{u} = (\frac{3}{5}, -\frac{4}{5})$.
4. $f(x_1, x_2) = (2x_1 - 3x_2)^2$, $\mathbf{u} = (0, 1)$.
5. $f(x_1, x_2) = e^{x_1 x_2 + 1}$, $\mathbf{u} = (-\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}})$.
6. $f(x_1, x_2) = \ln(x_1 + x_2)$, $\mathbf{u} = (-1, 0)$.

Hint: We have already found the partial derivatives in Exercise 5.

The directional derivatives are (in the same order):

1. $D_{\mathbf{u}}f(x_1, x_2) = \frac{1}{\sqrt{2}}(4x_1) + \frac{1}{\sqrt{2}}(-3) = 2\sqrt{2}x_1 - \frac{3}{\sqrt{2}}$.
2. $D_{\mathbf{u}}f(x_1, x_2) = -\frac{\sqrt{3}}{2}(2x_1 - x_2) + \frac{1}{2}(-x_1 + 2x_2) =$ $(-\sqrt{3} - \frac{1}{2})x_1 + (\frac{\sqrt{3}}{2} + 1)x_2$.
3. $D_{\mathbf{u}}f(x_1, x_2) = \frac{3}{5}(2x_1x_2^2 - 2x_2) - \frac{4}{5}(2x_1^2x_2 - 2x_1) =$ $\frac{6}{5}x_1x_2^2 - \frac{8}{5}x_1^2x_2 - \frac{6}{5}x_2 + \frac{8}{5}x_1$.
4. $D_{\mathbf{u}}f(x_1, x_2) = 0(8x_1 - 12x_2) + 1(18x_2 - 12x_1) = 18x_2 - 12x_1$.
5. $D_{\mathbf{u}}f(x_1, x_2) = -\frac{2}{\sqrt{5}}(x_2 \cdot e^{x_1x_2+1}) - \frac{1}{\sqrt{5}}(x_1 \cdot e^{x_1x_2+1}) =$ $(-\frac{1}{\sqrt{5}}x_1 - \frac{2}{\sqrt{5}}x_2) \cdot e^{x_1x_2+1}$.
6. $D_{\mathbf{u}}f(x_1, x_2) = -1(\frac{1}{x_1+x_2}) + 0(\frac{1}{x_1+x_2}) = -\frac{1}{x_1+x_2}$.

Optional: Some applications of gradient vectors

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

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

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

  where $\alpha > 0$. This is because the objective function at $\mathbf{x}$ increases 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

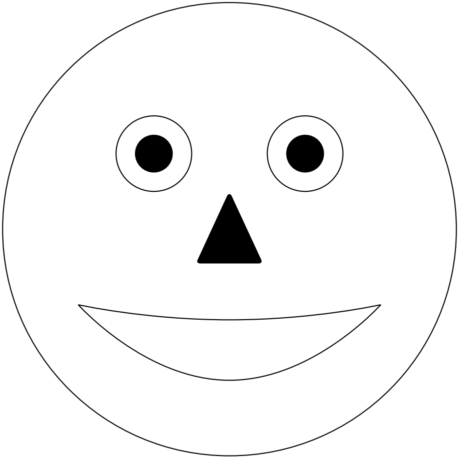
Some extra resources about functions and derivatives:

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

Q&A

Any questions?

Until the next time...



Thank you for your attention!