Functions

Functions

Recall the definition:

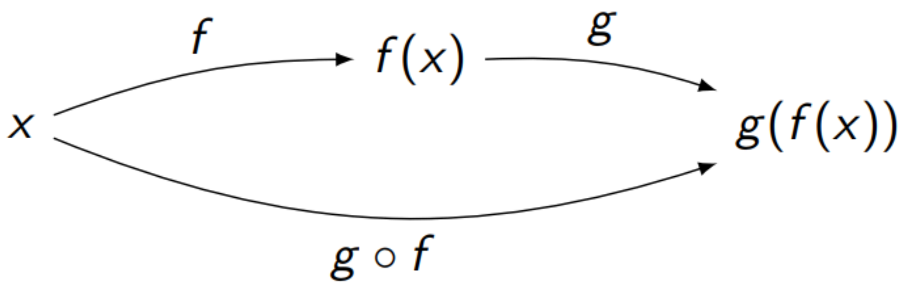
- Let A and B be sets of real numbers.
- A *function* $f : A- > B$ is a rule that maps each element in A to exactly one element in B.
- The unique element that x is mapped into, by the function f , is called the *image* of x under f, and is denoted $f(x)$.
- We will focus on functions that have a closed-form expression.
    - Some examples are
        - $f(x) = x^3 - 1, g(x) = sin(2x), h(x) = e^{-x^2}$

Compositions of functions

Definition 2 (Function composition)

- If $f : A- > B$ and $g : B- > C$,
- Then their *composition*,
    - $g \circ f(x) = A- > B$ (read "*g* of *f* "),
- Is defined as $g \circ f(x) = gf(x)$ wherever it exists.

Schematically:



- Compositions can be extended to three or more functions.

Exercise 1

Find $g \circ f(x)$ and $f \circ g(x)$:

1. $f(x) = x^2$ and $g(x) = cos(x)$
    a. $fg(x) = (cos(x))^2 = cos^2(x)$
    b. $gf(x) = cos(x^2)$
2. $f(x) = e^x$and $g(x) = x^3$
    a. $fg(x) = e^{(x^3)} = e^{x^3}$
    b. $gf(x) = (e^x)^3 = e^{x^3}$

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. $f(x) = 2^{x^3}$
    a. $g(x) = 2^x$
    b. $h(x) = x^3$
    c. $f(x) = gh(x)$
2. $f(x) = \frac{1}{sin(x)}$
    a. $g(x) = \frac{1}{x}$
    b. $h(x) = sin(x)$

c. $f(x) = gh(x)$

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

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. $f(\mathbf{x}) = 5$.

Answers:

- *Yes this function is linear*
- *Yes this function is linear*
- *No this function is not linear*
- *Yes this function is linear*

Sigmoid function

1. The *sigmoid function* $\sigma \rightarrow: R \rightarrow (0,1)$ is defined as
   a. $\sigma(x) = \frac{1}{1+e^{-x}}$
   b. The sigmoid function has the useful property of mapping any real number to the interval (0, 1) (prove it!)
      i. The function $\sigma(x) = \frac{1}{1+e^{-x}}$ is always positive because the denominator $1 + e^{-x}$ is always greater than 1.
      ii. As x ranges from negative infinity to positive infinity, $\sigma(x)$ asymptotically approaches 0 but never actually reaches it, and similarly approaches 1 but never actually reaches it.
      iii. Therefore, the output of the sigmoid function is always in the interval (0, 1).
2. $\sigma(x)$ is strictly increasing in *x* (prove it!).
   a. $\sigma'(x)$:
      i. Use chain rule:
         1. $f'(g(x)) * g'(x)$

ii. $f(x) = 1/y, g(x) = 1 + e^{-x}$

iii. $\sigma'(x) = -\frac{1}{(1+e^{-x})^2} * g'(x)$

iv. $\sigma'(x) = -\frac{1}{(1+e^{-x})^2} * -e^{-x}$

v. $\sigma'(x) = -\frac{-e^{-x}}{(1+e^{-x})^2}$

vi. $\sigma'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$

3. $\lim_{x\to-\infty} \sigma(x) = 0$ and $\lim_{x\to\infty} \sigma(x) = 1$ (prove it!).
   a. As $\sigma(x) = \frac{1}{1+e^{-x}}$ approaches ∞ e^{-x} tends to 0 which means $\sigma(x) = \frac{1}{1+e^{-x}}$ tends to $\frac{1}{1}$ = 1
   b. As $\sigma(x) = \frac{1}{1+e^{-x}}$ approaches $-\infty$ e^{-x} tends to infinity which means $1 + e^{-x}$ tends to infinity which means $\frac{1}{\infty}$ tends to 0
4. 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. $\sigma^{-1}(y) = \ln(\frac{y}{1-y})$.

- Proof of $\sigma(x) = 1 - \sigma(-x)$:
  - $\sigma(x) = \frac{1}{1+e^{-x}}$
    - $1 - \sigma(-x)$:
      - $1 - \frac{1}{1+e^{x}}$
      - $\frac{1+e^{x}}{1+e^{x}} - \frac{1}{1+e^{x}}$
      - $\frac{e^{x}}{1+e^{x}}$

        - $\frac{(e^x)}{(1+e^x)} * \frac{e^{-x}}{e^{-x}}$
        - $\frac{1}{(e^{-x}+1)}$
        - $\frac{1}{(1+e^{-x})}$
        - $\frac{1}{1+e^{-x}}$
        - $\sigma(x)$
- Proof of $\sigma'(x) = \sigma(x)(1-\sigma(x))$
    - $\sigma'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$
    - $\sigma(x) = \frac{1}{1+e^{-x}}$
        - $\sigma(x)(1-\sigma(x))$
            - $\frac{1}{1+e^{-x}}(1 - \frac{1}{1+e^{-x}})$
            - $\frac{1}{1+e^{-x}} - \frac{1}{(1+e^{-x})^2}$
            - $\frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2}$
            - $\frac{e^{-x}}{(1+e^{-x})^2}$
            - $\sigma'(x)$
- Proof of $\sigma^{-1}(y) = \ln(\frac{y}{1-y})$
    - $\sigma^{-1}(y)$:
        - $\sigma(x) = \frac{1}{1+e^{-x}}$
        - $y = \frac{1}{1+e^{-x}}$
        - $x = \frac{1}{1+e^{-y}}$
        - $x(1+e^{-y}) = 1$
        - $1 + e^{-y} = \frac{1}{x}$
        - $e^{-y} = \frac{1-x}{x}$
        - $e^y = \frac{x}{1-x}$
        - $y = \ln(\frac{x}{1-x})$
        - $\sigma^{-1}(x) = \ln(\frac{x}{1-x})$
        - $\sigma^{-1}(y) = \ln(\frac{y}{1-y})$

Derivatives from first principles

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.

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.

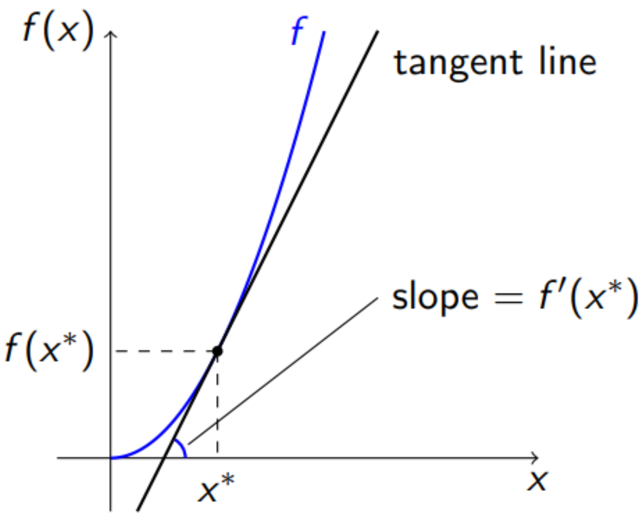


Figure 1: Slope of a function f at a point x^*.

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!

- 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

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

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

My own proof of the rest:

Exponential $c^x, c > 0$

$$f(x) = c^x$$

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

$$f'(x) = \lim_{t \to 0} \frac{c^x(c^t - 1)}{\triangle t}$$

$\frac{c^t - 1}{\triangle t}$ is a constant value therefore

$$ln(c) = \frac{c^t - 1}{\triangle t}$$

$$f'(x) = c^x ln(c)$$

Logarithmic $\log_c(x), 0 < c \neq 1$

$$f(x) = log_c(x)$$

$$f(x) = \frac{ln(x)}{ln(c)}$$

$$f'(x) = \frac{1}{xln(c)}$$

Sine $sin(x)$

$$f(x) = sin(x)$$

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

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

Using small angle approximations

$$\sin\theta \approx \theta$$
$$\cos\theta \approx 1 - \tfrac{1}{2}\theta^2$$
$$\tan\theta \approx \theta$$

$$f'(x) = \lim_{t \to 0} = \frac{sin(x)(1 - \frac{1}{2}t^2) + tcos(x) - sin(x)}{t}$$

$$f'(x) = \lim_{t \to 0} = \frac{sin(x) - \frac{1}{2}t^2 sin(x) + tcos(x) - sin(x)}{t}$$

$$f'(x) = \lim_{t \to 0} = \frac{-\frac{1}{2}t^2 sin(x) + tcos(x)}{t}$$

$$f'(x) = \lim_{t \to 0} = -\frac{1}{2}tsin(x) + cos(x)$$

$$f'(x) = \lim_{t \to 0} = cos(x)$$

Cosine $cos(x)$

$$f(x) = cos(x)$$

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

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

Using small angle approximations

$$\sin\theta \approx \theta$$

$$\cos\theta \approx 1 - \tfrac{1}{2}\theta^2$$

$$\tan\theta \approx \theta$$

$$f'(x) = \lim_{t \to 0} = \frac{cos(x)(1 - \frac{1}{2}t^2) - tsin(x) - cos(x)}{t}$$

$$f'(x) = \lim_{t \to 0} = \frac{cos(x) - \frac{1}{2}t^2cos(x) - tsin(x) - cos(x)}{t}$$

$$f'(x) = \lim_{t \to 0} = \frac{-\frac{1}{2}t^2cos(x) - tsin(x)}{t}$$

$$f'(x) = \lim_{t \to 0} = -\frac{1}{2}tcos(x) - sin(x)$$

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

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

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).

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

1. $f'(x) = -12x^2$
2. $f'(x) = 3x^2 + 2x$
3. $f'(x) = e^x + xe^x$
4. $f'(x) = \frac{xcos(x) - sin(x)}{x^2}$

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

Multivariate functions in general

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

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

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

6. $f'(x_1) = \frac{1}{x_1+x_2} / f'(x_2) = \frac{1}{x_1+x_2}$

Chain rule

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! □

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. $f(x_1, x_2) = 2x_2e^{x_1}$, where $x_1 = \ln(t^2+1)$, $x_2 = t^2$.
4. $f(x_1, x_2) = \sin(x_1x_2)$, where $x_1 = t$, $x_2 = \ln(t)$.

1. $\frac{df}{dt} = 4t^3 - 24t^5$
2. $\frac{df}{dt} = -2sin(x)cos(x) + 2sin(x)cos(x)$
3. $\frac{df}{dt} = 4te^{ln(t^2+1)} + \frac{4t^3}{t^2+1}e^{t^2+1} = 4t(t^2+1) + 4t^3 = 8t^3 + 4t$

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.

$g(w_0, w_1) = (f - y)^2$

To find the gradient vector of g with respect to $w = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}^T$, we need to compute the partial derivatives of gg with respect to w_0 and w_1.

First, let's rewrite g in terms of w_0 and w_1:

$g(w_0, w_1) = (w_0 + w_1 x - y)^2$

Now, let's calculate the partial derivatives:

1. Partial derivative with respect to w_0

$\frac{\sigma g}{\sigma w_0} = 2(w_0 + w_1 x - y)$

1. Partial derivative with respect to w_1:

$\frac{\sigma g}{\sigma w_1} = 2(w_0 + w_1 x - y)x$

$\frac{\sigma g}{\sigma w_1} = 2x(w_0 + w_1 x - y)$

Now, we can write the gradient vector:

$\frac{\sigma g}{\sigma w} = \begin{pmatrix} 2x(w_0 + w_1 x - y) \\ 2(w_0 + w_1 x - y) \end{pmatrix}$

Exercise 11

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

$f(w) = -ln(\frac{1}{1+e^{-(w_0+w_1 x)}}) = ln(1 + e^{-(w_0+w_1 x)})$

$f'(w_0) = -\frac{e^{-(w_0+w_1 x)}}{1+e^{-(w_0+w_1 x)}}$

$f'(w_1) = -\frac{w_1 e^{-(w_0+w_1 x)}}{1+e^{-(w_0+w_1 x)}}$

$f'(w) = \begin{pmatrix} -\frac{e^{-(w_0+w_1 x)}}{1+e^{-(w_0+w_1 x)}} \\ -\frac{w_1 e^{-(w_0+w_1 x)}}{1+e^{-(w_0+w_1 x)}} \end{pmatrix}$

$g(w) = -ln(1 - \frac{1}{1+e^{-(w_0+w_1 x)}}) = -ln(\frac{e^{-(w_0+w_1 xz)}}{1+e^{-(w_0+w_1 x)}}) = ln(\frac{1+e^{-(w_0+w_1 x)}}{e^{-(w_0+w_1 xz)}})$

$g(w) = ln(1 + e^{-(w_0+w_1 x)}) - ln(e^{-(w_0+w_1 x)})$

$g(w) = ln(1 + e^{-(w_0+w_1 x)}) + (w_0 + w_1 x)$

$$g'(w_0) = -\frac{e^{-(w_0+w_1x)}}{1+e^{-w_0+w_1x}} + 1$$
$$g'(w_1) = -\frac{xe^{-(w_0+w_1x)}}{1+e^{-w_0+w_1x}} + x$$

Some applications of gradient vectors

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.