## Gradient

In the case of **a univariate function**, it is simply the **first derivative at a selected point**. In the case of **a multivariate function**, it is a **vector of derivatives** in each main direction (along variable axes). Because we are interested only in a slope along one axis and we don’t care about others these derivatives are called **partial derivatives**.

A gradient for an n-dimensional function f(x) at a given point p is defined as follows:

Изображение выглядит как черный, снимок экрана, текст, черно-белый

Автоматически созданное описание

The upside-down triangle is a so-called *nabla* symbol and you read it “del”.

## Matrix differentiation

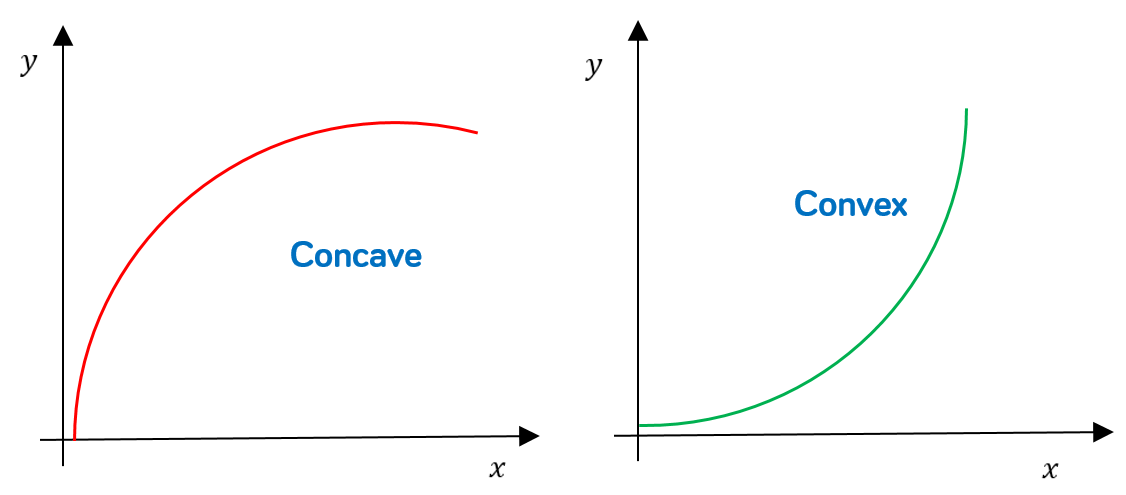
Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

## Convex and Concave functions

In [mathematics](https://en.wikipedia.org/wiki/Mathematics), a [real-valued function](https://en.wikipedia.org/wiki/Real-valued_function) is called **convex** if the [line segment](https://en.wikipedia.org/wiki/Line_segment) between any two distinct points on the [graph of the function](https://en.wikipedia.org/wiki/Graph_of_a_function) lies above the graph between the two points. Equivalently, a function is convex if its [*epigraph*](https://en.wikipedia.org/wiki/Epigraph_(mathematics)) (the set of points on or above the graph of the function) is a [convex set](https://en.wikipedia.org/wiki/Convex_set).

C**oncave function** is one for which the value at any convex combination of elements in the domain is greater than or equal to the convex combination of the values at the endpoints.



Let function f : RD → R be a function whose domain is a convex set. The function f is a convex function if for all x, y in the domain convex function of f, and for any scalar θ with 0 ⩽ θ ⩽ 1, we have

f(θx + (1 − θ)y) ⩽ θf(x) + (1 − θ)f(y).

A concave function is the negative of a convex function

A function f(x) is convex if and only if for any two points x, y it holds that

f(y) ⩾ f(x) + ∇xf(x) ⊤(y − x).

If we further know that a function f(x) is twice differentiable, that is, the Hessian exists for all values in the domain of x, then the function f(x) is convex if and only if ∇2 x f(x) is positive semidefinite

## Supporting hyperplane

Изображение выглядит как Красочность

Автоматически созданное описание

In [geometry](https://en.wikipedia.org/wiki/Geometry), a **supporting hyperplane** of a [set](https://en.wikipedia.org/wiki/Set_(mathematics)) S in [Euclidean space](https://en.wikipedia.org/wiki/Euclidean_space) Rn is a [hyperplane](https://en.wikipedia.org/wiki/Hyperplane) that has both of the following two properties:

* Sis entirely contained in one of the two [closed](https://en.wikipedia.org/wiki/Closed_set) [half-spaces](https://en.wikipedia.org/wiki/Half-space_(geometry)) bounded by the hyperplane,
* S has at least one boundary-point on the hyperplane.

## Legendre–Fenchel Transform and Convex Conjugate

The Legendre-Fenchel transform is a transformation from a convex differentiable function f(x) to a function that depends on the tangents s(x) = ∇f(x). It is worth stressing that this is a transformation of the function f(·) and not the variable x or the function evaluated at x. The Legendre-Fenchel transform is also known as the convex conjugate and is closely related to duality.

The convex conjugate of a function f : RD → R is a function f ∗ defined by

f ∗ (s) = supremum x∈RD (⟨s, x⟩ − f(x)) .

We can reconstruct any function f(x), with some restriction, by **just knowing its tangent line at each point on its graph**.

Describing the tangent line of a function, on the other hand, requires two pieces of information; the slope of the line at each point, given by the value of df/dx, and the y-interception of the line at each point, b.

Therefore, we can encode *all* the information of a function f(x) into just these two values and this is indeed *exactly* what the Legendre transformation does; **generates a new function from df/dx and b**.

The important thing about this is that the Legendre transformation of a function then contains exactly the same information as the original function, just “presented” in a different way. This is why it’s useful in the first place.

Изображение выглядит как текст, снимок экрана, Шрифт, дизайн

Автоматически созданное описание