Calculus Refresher (Derivatives)

Last Edited: February 21, 2022

ő A Note

This refresher will only cover concepts involving derivatives. For other topics in Calculus consult your favorite Calculus textbook.

We will also assume you are familiar with the definition of a function and limits.

ő Formal Definition of a Derivative

Suppose you have a function, $f : \mathbb{R} \to \mathbb{R}$, then the *derivative* of f with respect to x is defined as:

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \quad ,$$

provided this limit exists.

We often denote the derivative of f with respect to x as:

$$\frac{d}{dx} f \qquad \text{or} \qquad \frac{df}{dx} \qquad \text{or} \qquad f'(x) \; .$$

If f has a derivative we say it is *differentiable*.

ő Intuitive Understanding of a Derivative

One way to think of the derivative of f with respect to x is as a measure of the rate of change of f as you change x.

For example, if f denotes a position of someone, then $f'(x)$ would give the velocity of that same person at time x.

ő Common Derivatives

Polynomials

$$\frac{d}{dx}(c) = 0,\ \ c \in \mathbb{R}$$

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

Trigonometric Functions

$$\frac{d}{dx}(\sin(x)) = \cos(x)$$

$$\frac{d}{dx}(\cos(x)) = -\sin(x)$$

$$\frac{d}{dx}(\tan(x)) = \sec^2(x)$$

Exponential/Logarithm

$$\frac{d}{dx}(e^x) = e^x$$

$$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$$

ő Linearity of Differentiation

Suppose you have differentiable functions, $f : \mathbb{R} \to \mathbb{R}$ & $g : \mathbb{R} \to \mathbb{R}$, then:

$$\frac{d}{dx}\left(af(x) + bg(x)\right) = af'(x) + bg'(x),$$

$$a, b \in \mathbb{R}$$

Three Differentiation Rules

Suppose you have differentiable functions, $f : \mathbb{R} \to \mathbb{R}$ & $g : \mathbb{R} \to \mathbb{R}$ then:

<u>The Product Rule</u>

$$\frac{d}{dx}\left(f(x)g(x)\right) = f'(x)g(x) + f(x)g'(x)$$

<u>The Quotient Rule</u>

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{\left(g(x)\right)^2}$$, where $g(x) \neq 0$

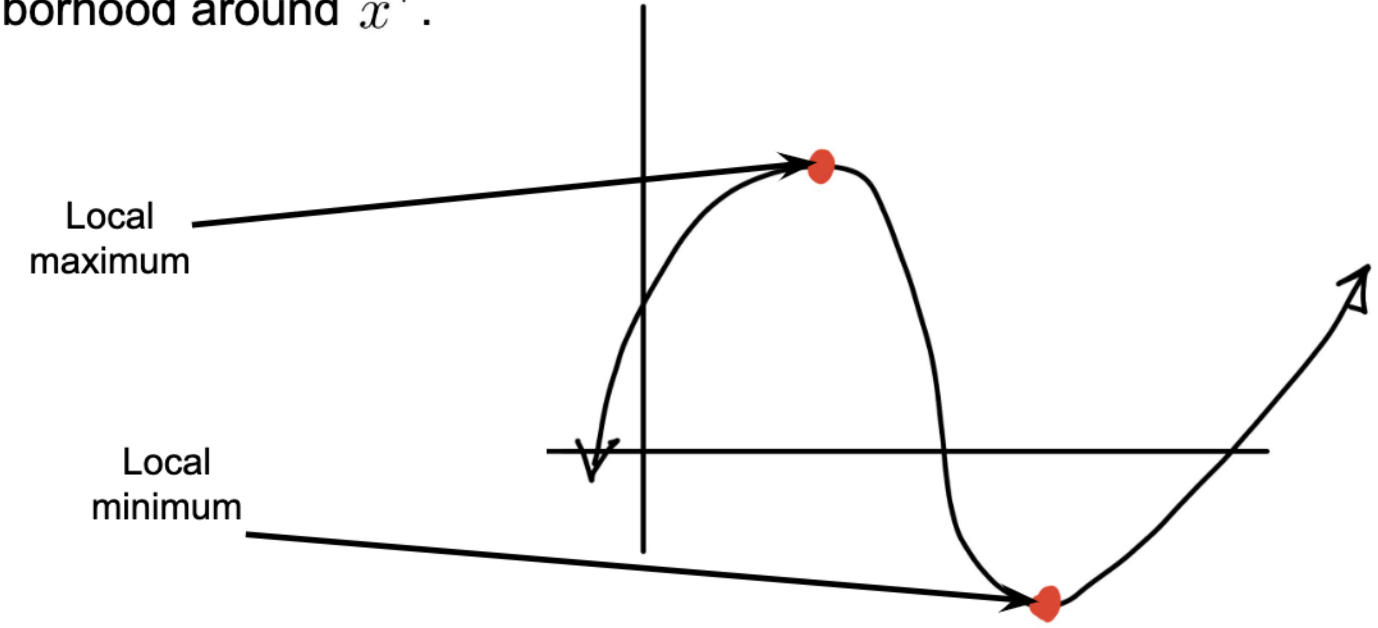
<u>The Chain Rule</u>

$$\frac{d}{dx}\left(f(g(x))\right) = f'(g(x))g'(x)$$

Local Maximums & Local Minimums

A function, f , is said to have a *local maximum* at x^* if $f(x^*) \geq f(x)$ for all x in some neighborhood around x^*.

A function, f , is said to have a *local minimum* at x^* if $f(x^*) \leq f(x)$ for all x in some neighborhood around x^*.



ő Fermat's Theorem

If a differentiable function, f , has a local minimum or maximum at x^* , then $f'(x^*) = 0$.

Note: that the converse is not necessarily true, meaning that $f'(x^*) = 0$ does not imply that f has a local minimum or maximum at x^*.

However, a strategy for finding a local maximum or minimum is to find where the derivative of f is 0.

ő Partial Derivatives

Suppose you have a function, $f : \mathbb{R}^n \to \mathbb{R}$, with inputs $x_1, x_2, \ldots, x_n$ the *partial derivative* of f with respect to x_i is:

$$\lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}.$$

Note that these are computed by treating the remaining input variables as constants and taking the single variable derivative of f with respect to x_i.

These are typically denoted as:

$$\frac{\partial}{\partial x_i} f \quad \text{or} \quad \frac{\partial f}{\partial x_i} \quad \text{or} \quad \partial_{x_i} f \quad \text{or} \quad f_{x_i}.$$

Intuitive Understanding of a Partial Derivative

One way to think of the partial derivative of f with respect to x_i is as a measure of the rate of change of f as you change x_i, while keeping all other inputs constant.

Gradients

Suppose you have a function, $f : \mathbb{R}^n \to \mathbb{R}$, with inputs $x_1, x_2, \ldots, x_n$ the *gradient* of f is the vector of partial derivatives of f,

$$\nabla f = (f_{x_1}, f_{x_2}, \ldots, f_{x_n}).$$

ő Gradients and Local Extrema

Suppose you have a differentiable function, $f : \mathbb{R}^n \to \mathbb{R}$. If f has a local maximum or minimum at x^*, then $\nabla f(x^*) = 0$.

Again, we note that $\nabla f(x^*) = 0$ does not imply that f has a local maximum or minimum at x^*. However, we often search for local extrema by finding where the gradient is equal to 0.

Gradient Descent

Suppose you have a differentiable function, $f : \mathbb{R}^n \to \mathbb{R}$.

It has been shown that the direction of the vector $\nabla f(x^*)$ is the direction you should "travel" from x^* in order to increase the value of f most rapidly.

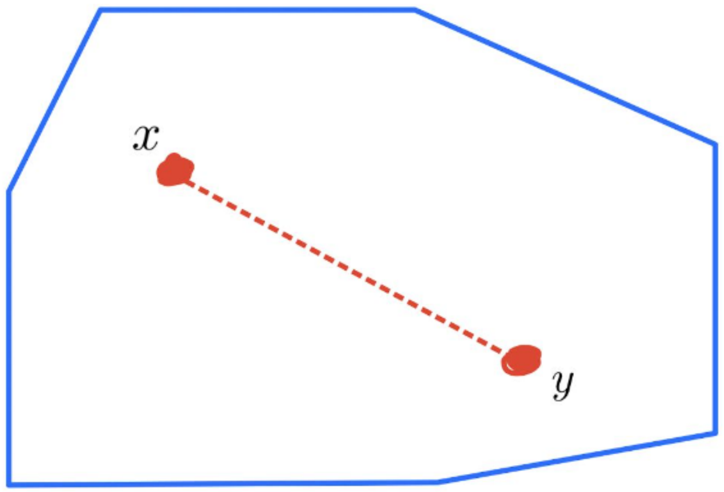
Conversely the direction of $-\nabla f(x^*)$ is the direction you should "travel" from x^* in order to decrease the value of f most rapidly.

Algorithms that leverage this latter fact in order to decrease a function are known as *gradient descent* methods.

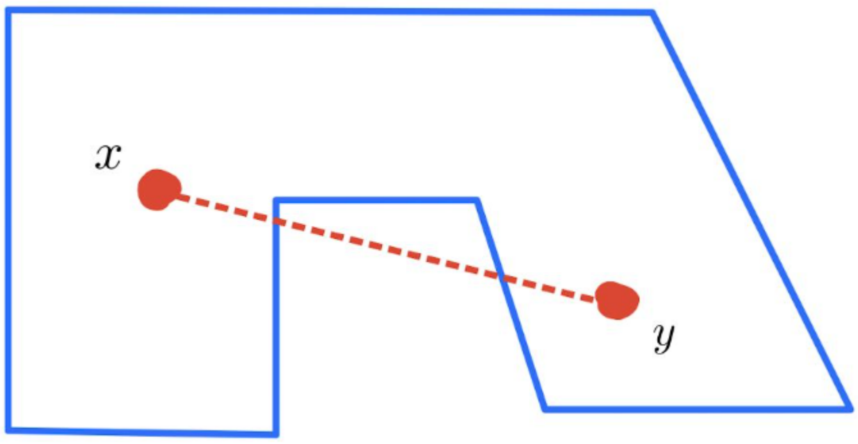
Convex Set

A set $\mathcal{S} \subset \mathbb{R}^n$ is *convex* if for $x, y \in \mathcal{S}$ and $\theta \in (0, 1)$ it holds that $\theta x + (1-\theta) y \in \mathcal{S}$.

Convex Set



Non-Convex Set



ő Convex Functions

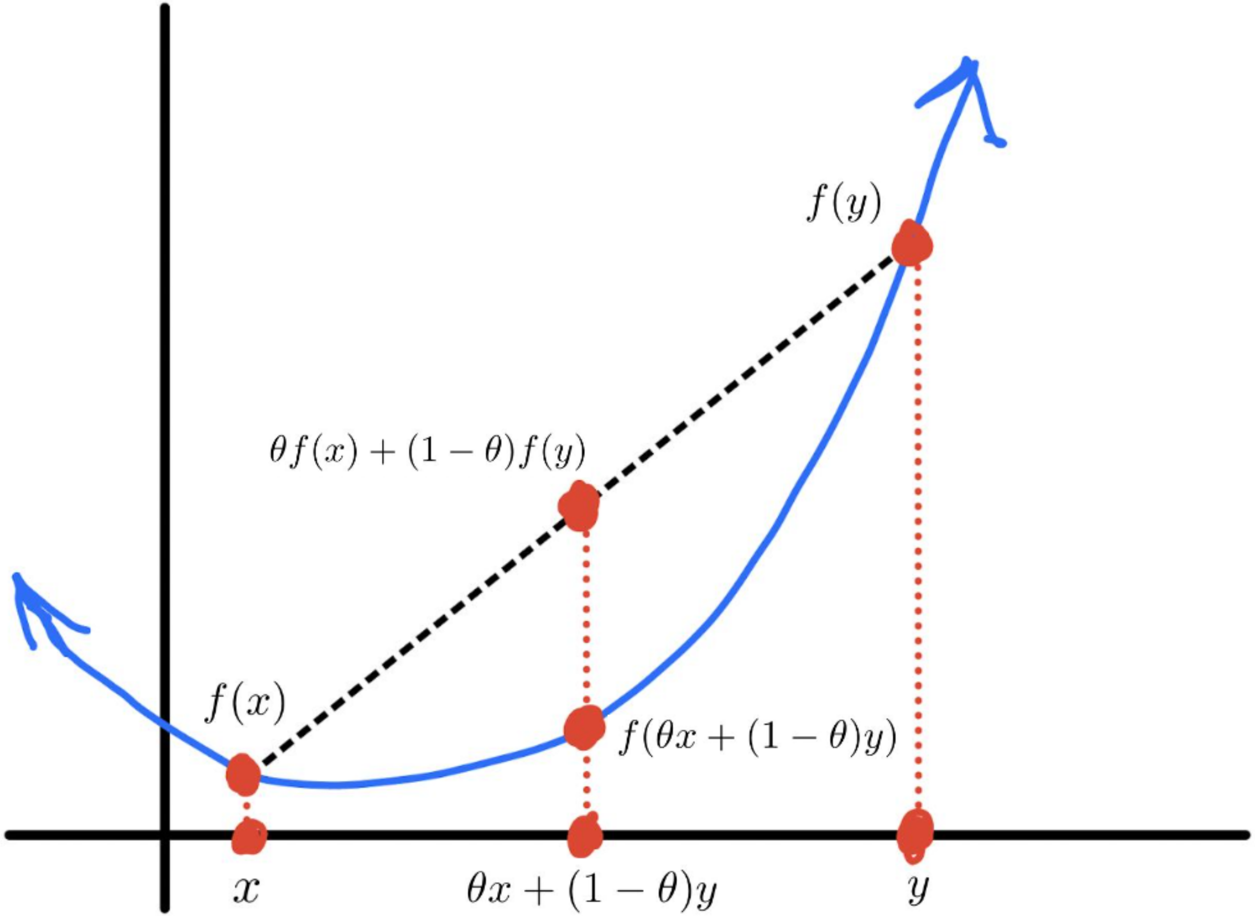
A function, $f : \mathbb{R}^n \to \mathbb{R}$, is *convex* if:

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

for $\theta \in (0, 1)$.

If the inequality is strict, we say that f is strictly convex.

ő Convex Function: Visualization



Ő Two Convex Function Facts

1. If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex set and $\mathcal{S}$ is a convex subset, then any local minimum of f is also a global minimum (in $\mathcal{S}$).
2. If $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex on a convex subset $\mathcal{S}$, then there exists at most one local minimum of f.

ő Learn More About Convex Optimization

Convex optimization is an import field for proving results in data science.

For the proofs of the two facts in the previous slide and to learn more about convex optimization see these pdfs.

- https://ai.stanford.edu/~gwthomas/notes/convexity.pdf
- https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf