Introduction to differentials

Outline

Introduction

Derivatives

The *derivative* of a function $f(x) : \mathbb{R} \to \mathbb{R}$ is defined to be

$$f'(x) \coloneqq \lim_{\delta \to 0} \frac{f(x+\delta) - f(x)}{\delta}.$$

- "If you make a very small change in x, how will f change?"
- And this is then made rigorous as the small change in x becomes small from the limit.

Leibniz notation

Intro calculus classes teach of Leibniz notation: $\mathrm{d}f/\mathrm{d}y$ as another way to write $f'(x)$:

$$\frac{\mathrm{d}f}{\mathrm{d}x} = f'(x) = \lim_{\delta \to 0} \frac{f(x+\delta) - f(x)}{\delta}$$

- Intuition: $\mathrm{d}x$ is an "infinitely small" change in x, $\mathrm{d}f$ is an "infinitely small" change in f
- Then taking the ratio gives the "instantaneous" rate of change of f: the derivative

We will ignore this way of thinking

- Instead, give *differentials* their own rigorous definition in general
- Then make connection to derivatives (if function is differentiable)

Partial derivatives

For a function $\mathbf{f}(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}^n$ (now multivariate), the *partial derivative* is

$$\frac{\partial f_i}{\partial x_j} := \lim_{\delta \to 0} \frac{f_i(\mathbf{x} + \delta \mathbf{e}_j) - f_i(\mathbf{x})}{\delta},$$

where $\mathbf{e}_j \in \mathbb{R}^m$ is the unit vector where the jth element is 1 and the rest are 0.

- Looking at how ith component of $\mathbf{f}$ changes with respect to jth component of $\mathbf{x}$.

Think of being acted on by the *differential operator* $\partial / \partial x_j$:

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial}{\partial x_j} f_i(\mathbf{x}),$$

we are "acting" on $f_i(\mathbf{x})$ by differentiating with respect to x_j

- *This operator is not a fraction! It is just the notation chosen to represent this action*

...

If $f(x) : \mathbb{R} \to \mathbb{R}$, then the partial derivative is

$$\frac{\partial f}{\partial x} = \lim_{\delta \to 0} \frac{f(x + \delta \cdot 1) - f(x)}{\delta} = f'(x).$$

The partial derivative is just... the derivative for univariate functions.

We will use the partial derivative notation when differentiating univarite functions

- Better for generalizing

Jacobian matrix

For a function $\mathbf{f}(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}^n$ put all partial derivatives into a matrix

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \begin{bmatrix} \partial f_1/\partial x_1 & \cdots & \partial f_1/\partial x_m \\ \vdots & \ddots & \vdots \\ \partial f_n/\partial x_1 & \cdots & \partial f_n/\partial x_m \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

This is commonly called:

- The *derivative* (understanding it is a combination of all partial derivatives).
- The *Jacobian* of $\mathbf{f}$ with respect to $\mathbf{x}$.
- Sometimes labeled as $\mathbf{J}_{\mathbf{x}}\mathbf{f}$ or $\mathbf{J}$

When we say "the derivative" of $\mathbf{f}(\mathbf{x})$, we are talking about the collection of all partial derivatives

Special cases of Jacobian

If $f(x) : \mathbb{R} \to \mathbb{R}$, then $\partial f / \partial x \in \mathbb{R}^{1 \times 1} = \mathbb{R}$, so it is just "the derivative"

If $f(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}$, then $\partial f / \partial \mathbf{x} \in \mathbb{R}^{1 \times m}$ is a *row vector*

- Aside: this is not the same as the *gradient*, $\nabla_{\mathbf{x}} f$, which is a column vector
- So $(\partial f / \partial \mathbf{x})^\top = \nabla_{\mathbf{x}} f$, they are transposes of each other

If $\mathbf{f}(x) : \mathbb{R} \to \mathbb{R}^n$, then $\partial \mathbf{f} / \partial x \in \mathbb{R}^{n \times 1} = \mathbb{R}^n$ is a *column vector*

Taylor series

For a function $f : \mathbb{R} \to \mathbb{R} \in \mathcal{C}^d$ (can be differentiated d times), it's *Taylor series centered at x_** is

$$f(x) = \sum_{i=0}^{d} \frac{1}{i!} \frac{\partial^i f(x_*)}{\partial x^i} (x - x_*)^i = f(x_*) + \frac{\partial f(x_*)}{\partial x} (x - x_*) + \cdots$$

For function $\mathbf{f}(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}^n$, Taylor series centered at $\mathbf{x}_*$ is

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_*) + \frac{\partial \mathbf{f}(\mathbf{x}_*)}{\partial \mathbf{x}} (\mathbf{x} - \mathbf{x}_*) + \cdots$$

- *We are mainly concered with the linear component of the Taylor series*
- Other terms of multivariate Taylor series are outside of scope

Differentials

Differentials

A *differential* for any function $f(x) : \mathbb{R} \to \mathbb{R}$ is defined as

$$\mathrm{d}f = \mathrm{d}f(x; \Delta) \coloneqq f(x + \Delta) - f(x).$$

Informally, is just the difference when you have a very small change in the input

- Δ is usually taken to be very small ($\Delta < 0.01$), *but $\Delta \neq 0$!*
- So the differential will depend on the value of Δ
- It is a function itself, $\mathrm{d}f : \mathbb{R} \to \mathbb{R}$

For special case $f(x) = x$,

$$\mathrm{d}f = (x + \Delta) - x = \Delta \coloneqq \mathrm{d}x.$$

The output is just the input, make this the differential of the input.

- Will write Δ as $\mathrm{d}x$ doing forward

Multivariate differentials

If $\mathbf{f}(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}^n$, a multivariate differential is just a vector:

$$\boldsymbol{\Delta} = \mathrm{d}\mathbf{x} = \begin{bmatrix} \mathrm{d}x_1 \\ \vdots \\ \mathrm{d}x_m \end{bmatrix}, \quad \mathrm{d}\mathbf{f} = \mathbf{f}(\mathbf{x} + \mathrm{d}\mathbf{x}) - \mathbf{f}(\mathbf{x}) = \begin{bmatrix} \mathrm{d}f_1 \\ \vdots \\ \mathrm{d}f_n \end{bmatrix}$$

Connection to derivatives

Relating $\mathrm{d}f$ to $\mathrm{d}x$

For a function $f : \mathbb{R} \to \mathbb{R}$, we can have differentials for the input ($\mathrm{d}x$) and output ($\mathrm{d}f$)

- *We would like to have a way to relate $\mathrm{d}f$ to $\mathrm{d}x$*
- "If we change input by Δ, how will output change?"
- This sounds a lot like a derivative... *but it does not need to be (e.g. Brownian motion)*

Univariate differentiable functions

If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at least once, the Taylor series of $f(x + \mathrm{d}x)$ centered at x is

$$f(x + \mathrm{d}x) = \sum_{i=1}^{\infty} \frac{1}{i!} \frac{\partial^i f(x)}{\partial x} (x + \mathrm{d}x - x)^i = f(x) + \frac{\partial f}{\partial x}\mathrm{d}x + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\mathrm{d}x^2 + \cdots .$$

Differential is then

$$\mathrm{d}f = f(x + \mathrm{d}x) - f(x) = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial^2 f}{\partial x^2}\frac{\mathrm{d}x^2}{2} + \cdots = \frac{\partial f}{\partial x}\mathrm{d}x + o(\mathrm{d}x).$$

As $\mathrm{d}x \to 0$, *the other terms all shrink much faster, so we ignore them*:

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x$$

This is the connection of differentials for a differentiable function

- Only looking at linear part of Taylor series
- Higher order terms technically still there, but are so small we can safely ignore
- *This will not always be the case (e.g. geometric brownian motion)*

Multivariate differentiable functions

Same idea with multivariate function $\mathbf{f}(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}^n$: Taylor series of $\mathbf{f}(\mathbf{x} + \mathrm{d}\mathbf{x})$ centered at $\mathbf{x}$ is

$$\mathbf{f}(\mathbf{x} + \mathrm{d}\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathrm{d}\mathbf{x} + \mathbf{o}(\mathrm{d}\mathbf{x}).$$

As $\mathrm{d}\mathbf{x} \to \mathbf{0}$, the $\mathbf{o}(\mathrm{d}\mathbf{x})$ can be ignored, resulting in

$$\mathrm{d}\mathbf{f} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathrm{d}\mathbf{x}$$

This works for any dimensionality:

- $f(x) : \mathbb{R} \to \mathbb{R}$ we already did
- $f(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}$ gives

$$\mathrm{d}f = \frac{\partial f}{\partial \mathbf{x}} \mathrm{d}\mathbf{x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_m} \end{bmatrix} \begin{bmatrix} \mathrm{d}x_1 \\ \vdots \\ \mathrm{d}x_m \end{bmatrix} = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j} \mathrm{d}x_j.$$

- $\mathbf{f}(x) : \mathbb{R} \to \mathbb{R}^n$, same idea