Cheatsheet

Contents

Calculus

How to learn

Remember this is a humble *cheatsheet*. It's only definitions. It's not useful to memorise definitions that can be found online. *You* need to develop intuition and understanding of all these definitions through the lectures and the 3blue1brown videos here.

While I find exercises really useful for most maths, I find them personally less useful for calculus intuition: calculating derivatives is very different from understanding them. However the notebooks are useful guides on how to implement calculus programatically, at the very least. Ideally, you should do the exercises on the 3blue1brown transcripts, the notebook, and the questions in this cheatsheet.

Calculus is foundational material that pops up again and again. Just like linear algebra, each time you learn it, you get better intuition into what's going on. Beginners won't learn as much as people who've seen it before, and that's OK.

- Ideally, a calculus aficionado would still have gone through everything on the 3blue1brown course. Even if you've done Calculus before, you'll get more insight by going over the geometric approach he uses carefully.

- If you're experienced and can do that, go through Chapters 1, 2, and 3 here, instead of doing the notebooks.

For the 3blue1brown link, you can skip the videos/ transcripts on:

- (ε, δ definitions of limits)
- L'hopitals rule
- Integration
- The other way to visualise derivatives
- Divergence and Curl

… however they are all massively useful so go through if you have time!

What's the point?

Calculus is often described as the mathematics of how things *change*. It's fundamental to all aspects of physics and engineering, or at least those that deal with systems that change in time or space.

But apart from intrinsic beauty and interest, why does a data scientist or an AI specialist need calculus?…

Machine learning algorithms all go through a *training step*. This is where they *change* their behaviour based on receiving new data. The

old-school word for training/learning is *optimisation* (which I prefer, but isn't as good at attracting venture capital).

- Understanding (some aspects of) calculus is necessary to understanding how training works at a conceptual level, across algorithms.
- Understanding *automatic differentiation* is necessary to understanding how training is implemented at a practical level.

But even apart from that, calculus just crops up in so many places whenever you deal with maths that it's a shame to not have some understanding of it.

Note that traditionally, calculus courses divide into two sections: *differentiation* and *integration*. We will not do integration due to time constraints and irrelevance to the above goals 😔.

Notation for changing quantities

	Description	Notes
x, y	Arbitrary variable names	
t	Specific variable name: t usually denotes time	
Δx, Δy	Traditionally denotes a *change* in the variable	Δ is the capitalised greek letter Delta.
$\mathrm{d}x$, $\mathrm{d}y$	*Tiny* changes in x and y. Often called *infinitesimal* changes.	See Section 1.2 for how to do algebra with these
$f(x)$	f is a function. x is the *independent* variable $f(x)$ is the *dependent* variable	Independent variables are inputs which change of their own accord. Dependent variables change only due to the independent vairable
$x(t)$	x is a function. t, the independent variable, is time.	Students are used to x being an independent variable, as above, so get confused

1. Understanding derivatives through limits

What is a derivative? A fully rigorous definition requires an entire semester of "Real Analysis" (a topic in maths). A reasonably rigorous definition would require, at a minimum:

- The (ε, δ) definitions of limits of sequences and functions
- The (ε, δ) definition of function continuity

You're optionally welcome to learn these by taking e.g. this Khan academy course. If you're really ambitious and have time, take a course in real analysis! Little Rudin is a difficult but rewarding book on proving calculus from first principles. But that's not the goal of this course or cheatsheet!

I'm interested in giving you an *intuitive* understanding of calculus that's useful in an applied setting. I believe this is possible without the rigorous understanding. I also believe it makes learning calculus from first principles easier, if this is something you ever want to do.

I certainly didn't get a *feel* for calculus by learning it the rigorous way!

So we start on the first goal, defining a derivative in a way that's rigorous enough to make sense and be consistent, but doesn't take an entire semester.

Differentiation is the process of finding a derivative

1.1. Case study: position and velocity

Pause a video at an instant in time. Nothing has a velocity if the video is paused: nothing is moving. Velocity doesn't make sense at an instant in time because it is a *differential quantity*.

Defined in Definition 1.2.1

Velocity only makes sense by *comparing* positions at two (close-together) timepoints. It's a *comparison* of position against time. You have to play the video for as short a time as your finger allows, and see how much the characters have moved to infer their velocity. As such, at school, you're told that

$$\text{velocity} = \frac{\text{change in position}}{\text{change in time}} = \frac{\Delta \text{ position}}{\Delta \text{ time}} \tag{1}$$

Let's denote your position as a function $x(t)$. Then your *average* velocity over the time interval $[t, t + \Delta t]$ is

$$\frac{x(t+\Delta t)-x(t)}{\Delta t} \tag{2}$$

This is an inaccurate average if Δt is large. You moved 40 miles in your car in an hour , but this isn't reflective of your velocity at time 39:52.

So $x(t+\Delta t)-x(t) = 40$ miles, and $\Delta t = 1$hour

How do we figure out the velocity at an instant in time? We shrink the time interval Δt in Equation 2. As we shrink it further (from an hour to a minute to a second), the accuracy of the velocity increases. Mathematically, we write this as:

$$\lim_{\Delta t \to 0} \frac{x(t+\Delta t)-x(t)}{\Delta t} \tag{3}$$

Now if we set $\Delta t = 0$, we get $\frac{0}{0}$ which is undefined. But as Δt gets *closer* to zero, the fraction gets ever closer to a number that *is* defined. Like 4. This is the fundamental weirdness of calculus that's important to appreciate. Let's show by an explicit example, where $x(t) = t^2$:

Explicit calculation of velocity

Suppose $x(t) = t^2$. So your position is changing faster and faster as time goes on! What is your velocity at an arbitrary timepoint t?

$$\begin{aligned} x(t+\Delta t)-x(t) &= (t+\Delta t)^2 - t^2 \\ &= t^2 + 2t\Delta t + (\Delta t)^2 - t^2 \\ &= 2t\Delta t + \Delta t^2 \end{aligned} \tag{4}$$

Now as we shrink the time interval Δt to 0, the above equation (change in position) will also shrink to zero.

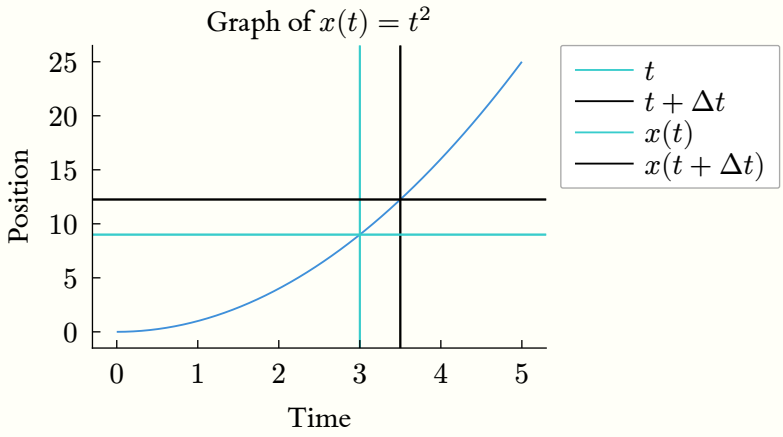
However…! the *ratio* of change in position to change in time, will not hit zero!

$$\begin{aligned} \frac{x(t+\Delta t)-x(t)}{\Delta t} &= \frac{2t\Delta t + (\Delta t)^2}{\Delta t} \\ &= 2t + \frac{(\Delta t)^2}{\Delta t} \end{aligned} \tag{5}$$

See? We get a term that is *independent* of Δt. So

$$\begin{aligned} \lim_{\Delta t \to 0} \frac{x(t+\Delta t)-x(t)}{\Delta t} &= \lim_{\Delta t \to 0} 2t + \Delta t \\ &= 2t \end{aligned} \tag{6}$$

Make sure you understand and can do this calculation before going on



Movement of my hand over time during a lecture. The velocity at time t is the *slope* of the curve at t, as you can see by shrinking Δt on the diagram. Make sure you get this graphical intuition, from the lectures or online videos.

Potential confusion

Look at the graph above, of $x(t) = t^2$. Notice that when we write $x(t) = t^2$, t is an arbitrary input to the function $x(t)$.

However… we then choose a particular timepoint t (position of vertical line). This is a different t. It is a fixed number, e.g. $t = 3$

The example above is *weird*.

Set $\Delta t = 0$	Set $\Delta t \neq 0$
Equation 3 is undefined	Equation 3 equals $2t + \Delta t$

In other words, the most accurate approximation of the velocity *now* is to compare positions at timepoints over a tiny tiny time interval. But the interval can't be zero, as change in position over 0 time will be 0.

1.2. Computing with infinitesimal numbers

To deal with this weirdiness, we use *infinitesimal* numbers. The word infinitesimal is unnecessarily loaded. Really, just imagine they are tiny but concrete, nonzero numbers, like -0.0000001 We

prepend a variable name with the letter 'd' to denote an infinitesimal change. So for our example of $x(t) = t^2$:

Before	$\Delta x(t) = 2t\Delta t + (\Delta t)^2$
After	$\mathrm{d}x(t) = 2t\,\mathrm{d}t + (\mathrm{d}t)^2$
Divide through by d*t*	$\frac{\mathrm{d}x}{\mathrm{d}t}(t) = 2t + \mathrm{d}t$
Ignore tiny terms	$\frac{\mathrm{d}x}{\mathrm{d}t}(t) = 2t$

Notice that we treated $\mathrm{d}x(t)$ as a *function* (hence the round brackets with t as an input). The change in position *depends* on the current time. We treat position as a *dependent* variable, and time as an *independent* variable. This means that $\frac{\mathrm{d}x}{\mathrm{d}t}(t)$, the velocity, is also treated as a function of t.

Notice that we killed $\mathrm{d}t$, because it's infinitesimal. But we *didn't* kill $\frac{\mathrm{d}x}{\mathrm{d}t}$, which is the *ratio* of infinitesimals. This is because a tiny number divided by another tiny number can be quite big! What is $\frac{0.00001}{0.0000001}$?

More generally, we have particular rules for when we do/don't ignore infinitesimal numbers.

Properties of infinitesimal numbers

Operation	Example	Action
Ratio	$\frac{\mathrm{d}x}{\mathrm{d}t}$	*Don't ignore*. $\frac{\text{tiny}}{\text{tiny}}$ might be big
Equate with normal numbers	$y = 2 + \mathrm{d}x$	*Ignore*. LHS is normal-sized so $\mathrm{d}x$ is negligible in comparison
Equate with infinitesimal numbers	$\mathrm{d}y = 2\,\mathrm{d}x$	*Don't ignore*. Both sides of equation are tiny, so tiny quantities are relevant
Equate with normal numbers	$\mathrm{d}y = 2 + \mathrm{d}x$	*You've made a mistake*. The LHS is infinitesimal but the RHS is not.
Ratio with exponents	$y = \frac{(\mathrm{d}x)^2}{\mathrm{d}t} + 3$	*Ignore*. $\frac{\text{tiny}^2}{\text{tiny}} = \text{tiny}\ \left(\frac{\text{tiny}}{\text{tiny}}\right) = \text{tiny}(\text{normal}) = \text{tiny}$
Ratio of normal to tiny equated with normal	$y = \frac{4x}{\mathrm{d}t}$	*You've probably made a mistake*. $\frac{\text{normal}}{\text{tiny}}$ is infinitely large.

The golden rule above is to figure out the exponent of tiny terms on each side of the equation . The side of the equation with the *lowest* max exponent retains all its terms. Any exponents that are *higher*, on the other side of the equation, are ignored.

(e.g. $\mathrm{d}x\,\mathrm{d}y$ and $(\mathrm{d}t)^2$ have exponent two as they are two tiny terms multiplied)

Example

$\mathrm{d}y(4\,\mathrm{d}x + \mathrm{d}x\,\mathrm{d}t) = (\mathrm{d}t)^2$

- Notice the LHS has a term with exponent 3 ($\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}t$), whereas the RHS has highest exponent 2. So we're going to kill any exponent 3 terms.

why? Let's simplify to make this clear:

$$\frac{\mathrm{d}y}{\mathrm{d}t}\left(4\frac{\mathrm{d}x}{\mathrm{d}t}\right) + \frac{\mathrm{d}y}{\mathrm{d}t}\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t = 1 \tag{7}$$

Notice that each fraction is a potentially non-tiny quantity. However, two fractions multiplied by $\mathrm{d}t$ are tiny, so we kill them and get

$$\frac{\mathrm{d}y}{\mathrm{d}t}\left(4\frac{\mathrm{d}x}{\mathrm{d}t}\right) = 1 \tag{8}$$

Another way of thinking about this: we are treating infinitesimal numbers as sufficiently small that taking exponents (powers) is much more important than adding / subtracting / (multiplying with non-infinitesimal numbers). $(\mathrm{d}x)^2$ is incomparably smaller than $0.0001\,\mathrm{d}x$ because 0.00001 is a normal (albeit small) number, which is incomparably bigger than $\mathrm{d}x$

Definition 1.2.1: Differential quantities

A differential quantity is a quantity that is a ratio of infinitesimal quantities. Velocity is the obvious example:

$$\begin{aligned}\text{Velocity}(t) &= \frac{\text{infinitesimal change in position}(t)}{\text{infinitesimal change in time}} \\ &:= \frac{\mathrm{d}x}{\mathrm{d}t}(t)\end{aligned} \tag{9}$$

"Calculus is all about comparing how fast different infinitesimal quantities change with respect to each other."

1.3. Terminology

1.3.1. Derivatives

Let $y(t) = 3t + 4$ denote the horizontal velocity of a meteor in space. Let $x(t)$ denote its position.

Then, given the previous sections, you should know that

$$y(t) = \frac{\mathrm{d}x}{\mathrm{d}t}(t) \tag{10}$$

We say that $y(t)$ *is the derivative of* $x(t)$*, with respect to (the independent variable)* t.

Definition 1.3.1.1: The derivative of a function

Suppose $f : \mathbb{R} \to \mathbb{R}$ is a function of an independent variable x.

Then we can make a new function: $\frac{\mathrm{d}f}{\mathrm{d}x} : \mathbb{R} \to \mathbb{R}$ which takes the same variable x as an input, and returns $\frac{\mathrm{d}f}{\mathrm{d}x}$ at x as the output.

We call $\frac{\mathrm{d}f}{\mathrm{d}x}$ the *derivative* of f with respect to x.

Notice the difference between f or $\frac{\mathrm{d}x}{\mathrm{d}x}$, which are functions, and $f(x)$ or $\frac{\mathrm{d}f}{\mathrm{d}x}(x)$, which are the *outputs* of the same functions given an input x.

1.3.2. Higher order derivatives

Let's get back to our velocity function: $y(t) = 3t + 4$. We can differentiate $y(t)$ again with respect to t!:

$$\frac{\mathrm{d}y}{\mathrm{d}t}(t) := \frac{\mathrm{d}^2x}{\mathrm{d}t}(t)$$

$\frac{\mathrm{d}y}{\mathrm{d}t}(t)$ represents $\frac{\text{infinitesimal change in velocity}}{\text{infinitesimal change in time}}$... so it's *acceleration*! It is the *double* derivative of position. A notation for the double derivative is in the equation above.

"Velocity is to position is as as acceleration is to velocity"

Notations for derivatives

Notation	Notes
$\frac{\mathrm{d}f}{\mathrm{d}x}(x)$	The d is `\mathrm{d}` in LaTeX.
$\frac{\mathrm{d}^2f}{\mathrm{d}x^2}(x)$	The derivative of the derivative. Also a function.
$f'(x) := \frac{\mathrm{d}f}{\mathrm{d}x}(x)$	"*Prime notation*". Less clear but common notation for the derivative. The independent variable is x
$\dot{x}(t) := \frac{\mathrm{d}x}{\mathrm{d}t}(t)$	"*Dot notation*". First derivative of function $x(t)$. Commonly used when independent variable is time.
$f''(x) = \frac{\mathrm{d}^2f}{\mathrm{d}x^2}(x)$	Prime notation for double derivative
$\ddot{x}(t) = \frac{\mathrm{d}^2x}{\mathrm{d}t^2}(t)$	Dot notation for double derivative
$f''''(x), \frac{\mathrm{d}^nf}{\mathrm{d}x^n}, \dddot{x}(t)$	Higher order derivatives

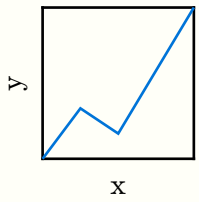
Question

Let position of an object at time t be described by the function $x(t) = t^2 + 4$

1. Modify the working in the previous section to find out that $\frac{\mathrm{d}x}{\mathrm{d}t}(t) = 2t$
2. Do your own working to show that the acceleration function is $\frac{\mathrm{d}^2x}{\mathrm{d}t^2}(t) = 2$

1.4. When can you compute a derivative

Derivatives are not always defined! IE it's physically impossible to differentiate some functions.



Why are derivatives undefined at sharp corners? Compare the effect on the dependent variable of an infinitesimal *negative* change in the independent variable vs a positive change.

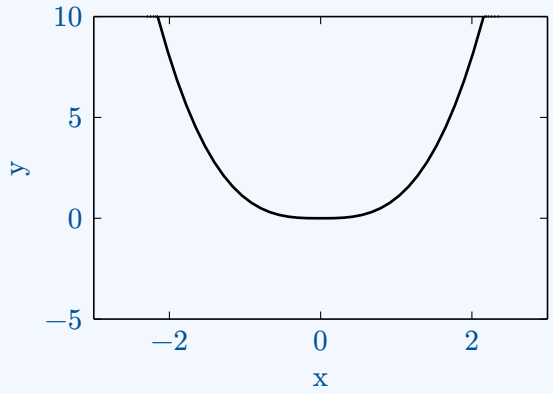
There is lots of maths and complication on when you can/not define a derivative. It boils down to:

"Derivatives are undefined when the function has corners or discontinuities"

Question

Here is a plot of

$$f(x) = \begin{cases} -x^3 \text{ if } x < 0 \\ x^3 \text{ if } x > 0 \end{cases} \tag{11}$$



Note that the derivative of $g(x) = x^3$ is $\frac{\mathrm{d}g}{\mathrm{d}x}(x) = 3x^2$

- Why is $f(x)$ once differentiable, but not twice differentiable?

1.5. Computing derivatives

- We'll get practice on doing this on computers in the notebook.
- We won't go through the rule for calculating derivatives of simple functions by hand ("*symbolic differentiation*"). I don't find it conceptually or practically useful.

The skill you *really* need is to sketch derivatives from graphs of a function $f(x)$. We practiced this in the lecture. Here is a nice quiz on practicing these sketches.

Common derivatives

Function $f(x)$	Derivative $\frac{\mathrm{d}f}{\mathrm{d}x}$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$x^n, n \in \mathbb{Z}$	nx^{n-1}
$\exp(x)$	$\exp(x)$
$\ln(x)$	$\frac{1}{x}$

e.g. the derivative of x^5 is $5x^4$

1.5.1. Antiderivatives / integrals

- Let $y(t) = 3t + 4$ denote the horizontal velocity of a meteor in space.
- Let $x(t)$ denote its position

How would we get a formula for $x(t)$ from the formula for $y(t)$? We have to *undo* taking the derivative. This is called *antidifferentiation*, or *integration*. In real life (unlike textbooks), this is *much harder* than differentiation. We won't learn about it in this course.

2. Differential operators

Notice in previous sections that we took the derivative of a *function*. And the outcome was also a *function*. For example, position was a *function* of time, and differentiating position yielded velocity, which was also a *function* of time.

As such, we can think of differentation as an *operator*, a mapping that takes in a function, and spits out a function.

Different types of functions: terminology

Object	Example	Description
Function (between scalars/vectors)	$f : \mathbb{R}^n \to \mathbb{R}^n$ $f(x) = (x^T x)Ax$	Input is a number/vector. So is output $y = f(x)$ *Don't* confuse f (function) with y (output)
Operator (between functions)	$\mathrm{D}_x(f) = \frac{\mathrm{d}f}{\mathrm{d}x}$	Since the output is a function, we need to evaluate it to get a number: $\mathrm{D}_x(f)(x)$
$x(t)$	x is a function. t, the independent variable, is time.	Students are used to x being an independent variable, as above, so get confused

2.1. Linearity of the differential operator

Look back on your linear algebra notes, and recall the definition of linearity.

Theorem 2.1.1 : Linearity of the differential operator

Take any differentiable functions $f : \mathbb{R} \to \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$. Then

$$\mathrm{D}_x(af + bg) = a\mathrm{D}_{x(f)} + b\mathrm{D}_x(g) \qquad \forall a, b \in \mathbb{R} \tag{12}$$

Am steadily making my theorems more mathematical and less *Englishy*.

We won't prove this theorem in the course, although if you've taken courses on limits, then it's not too difficult.

Question

Let

$f(x) = \sin(x)$ $\quad$ $g(x) = \cos(x)$

Note that

$\frac{\mathrm{d}f}{\mathrm{d}x}(x) = g(x)$ $\quad$ $\frac{\mathrm{d}g}{\mathrm{d}x}(x) = -f(x)$

Now let $h(x) = 2\cos(x) + 3\sin(x)$

1. Use linearity of the differential operator to calculate $\mathrm{D}_x(h)$ from the derivatives for f and g

2. Is it possible, or not, to calculate the derivative of $j(x) = \cos(4x) + \sin(5x)$ in the same way?

2.1.1. Understanding the differential operator on polynomials.

Let's consider degree 3 polynomials (cubic equations). An example is $f(x) = -x^3 + 4x - 6$.

We can represent degree 3 polynomials as a 4 dimensional vector space, whose bases are

$\underline{e}_1 = 1$ $\quad$ $\underline{e}_2 = x$

$\underline{e}_3 = x^2$ $\quad$ $\underline{e}_4 = x^3$

For instance, in this basis, we have $f = \begin{bmatrix} -1 \\ 0 \\ 4 \\ -6 \end{bmatrix}$

Question

1. Make sure you understand why degree 3 polynomials form a vector space, and the bases above span the vector space.

Remember from linear algebra that *every* linear map on a vector space can be represented as a matrix.

2. Figure out which matrix represents the differential operator on this vector space. To start, notice that when you apply this matrix to the vector representation of f, you should get

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \begin{bmatrix} 4 \\ 0 \\ -3 \\ 0 \end{bmatrix} \tag{13}$$

i.e. $\frac{\mathrm{d}f}{\mathrm{d}x}(x) = 0x^3 - 3x^2 + 0x + 4$

3. How would you use this matrix to differentiate a polynomial twice?
4. What would the differential operator on degree n polynomials look like?

The above question is leading because any infinitely differentiable function can be represented as a polynomial of `infinite` degree. That is, 1, x, x^2, ... form an infinite basis for this vector space of functions. For instance, here is a polynomial representation of the sin function.

Take-home message: you can think of the differential operator as an infinite dimensional square matrix.

2.2. Differential operator on products of functions

Theorem 2.2.1: The product rule

Take any differentiable functions $f : \mathbb{R} \to \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$. Then

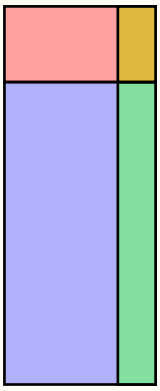
$$\begin{gathered} \mathrm{D}_x(fg) = f\frac{\mathrm{d}g}{\mathrm{d}x} + g\frac{\mathrm{d}f}{\mathrm{d}x} \\ \forall a, b \in \mathbb{R} \end{gathered} \tag{14}$$

Example

- The derivative of $f(x) = x$ is $\frac{\mathrm{d}f}{\mathrm{d}x}(x) = 1$. You should be able to prove this for yourself.
- We showed with our position-velocity example earlier that the derivative of $g(x) = x^2$ is $\frac{\mathrm{d}g}{\mathrm{d}x}(x) = 2x$

We can use the product rule to find the derivative of $h(x) = f(x)g(x) = x^3$

$$\begin{aligned} \frac{\mathrm{d}h}{\mathrm{d}x}(x) &= \frac{\mathrm{d}f}{\mathrm{d}x}(x)g(x) + \frac{\mathrm{d}g}{\mathrm{d}x}(x)f(x) \\ &= (1)x^2 + (2x)(x) \\ &= 3x^2 \end{aligned} \tag{15}$$



Label the sides of the axes on this diagram and use them to explain the product rule to yourself. Use the lectures if you're stuck. The largest rectangle should have area $f(x)g(x)$

2.3. Differential operator on compositions of functions

Theorem 2.3.1: The chain rule

Take any differentiable functions $f : \mathbb{R} \to \mathbb{R}$, and $g : \mathbb{R} \to \mathbb{R}$. Then

$$\mathrm{D}_x(f \circ g) = \mathrm{D}_g(f) \times \mathrm{D}_x(g) \tag{16}$$

Normally wouldn't use the $\times$ symbol for multiplication, but makes it visually easier here. It is just normal multiplication

This can't be easily visualised on a static page. Instead, go to the 3blue1brown video here

What does D_g mean? The subscript on the differential operator denotes the variable with which it is differentiating with respect to. What does it mean to differentiate with respect to g? Let's do this by example.

Example

- The derivative of $f(x) = x^2$ is $\frac{\mathrm{d}f}{\mathrm{d}x}(x) = 2x$..
- The derivative of $g(x) = \cos(x)$ is $\frac{\mathrm{d}g}{\mathrm{d}x}(x) = -\sin(x)$

We can use the chain rule to find the derivative of $h(x) = f(x) \circ g(x) = (\cos(x))^2$

1. $\mathrm{D}_g(f)$ is $2g$ as $f(g) = g^2$
2. $\mathrm{D}_x(g)$ is $-\sin(x)$

Overall

$$\begin{aligned}\frac{\mathrm{d}h}{\mathrm{d}x}(x) &= 2g \times -\sin(x) \\ &= 2\cos(x) \times -\sin(x) \\ &= -2\sin(x)\cos(x)\end{aligned} \tag{17}$$

3. Summary

This cheatsheet was all about making *algebras* on different objects. Rules for how they should interact and make babies. In particular:

We defined an algebra on *infinitesimal numbers*. You should know when to ignore them.

We already have an algebra on functions: you can add, scale, multiply, and compose them. We used this to make an algebra on the *differential operator*: how to add, scale, multiply and compose the *derivative* of functions. In particular,

- Adding and scaling were easy, as the differential operator is linear.
- Multiplying and composing were harder: we needed to build the product and chain rules.

In later parts of the course, we will go over *use cases*. A particular way of calculating derivatives, called *automatic differentiation*, underlies all training in machine learning. This notebook preps you for understanding autodiff.