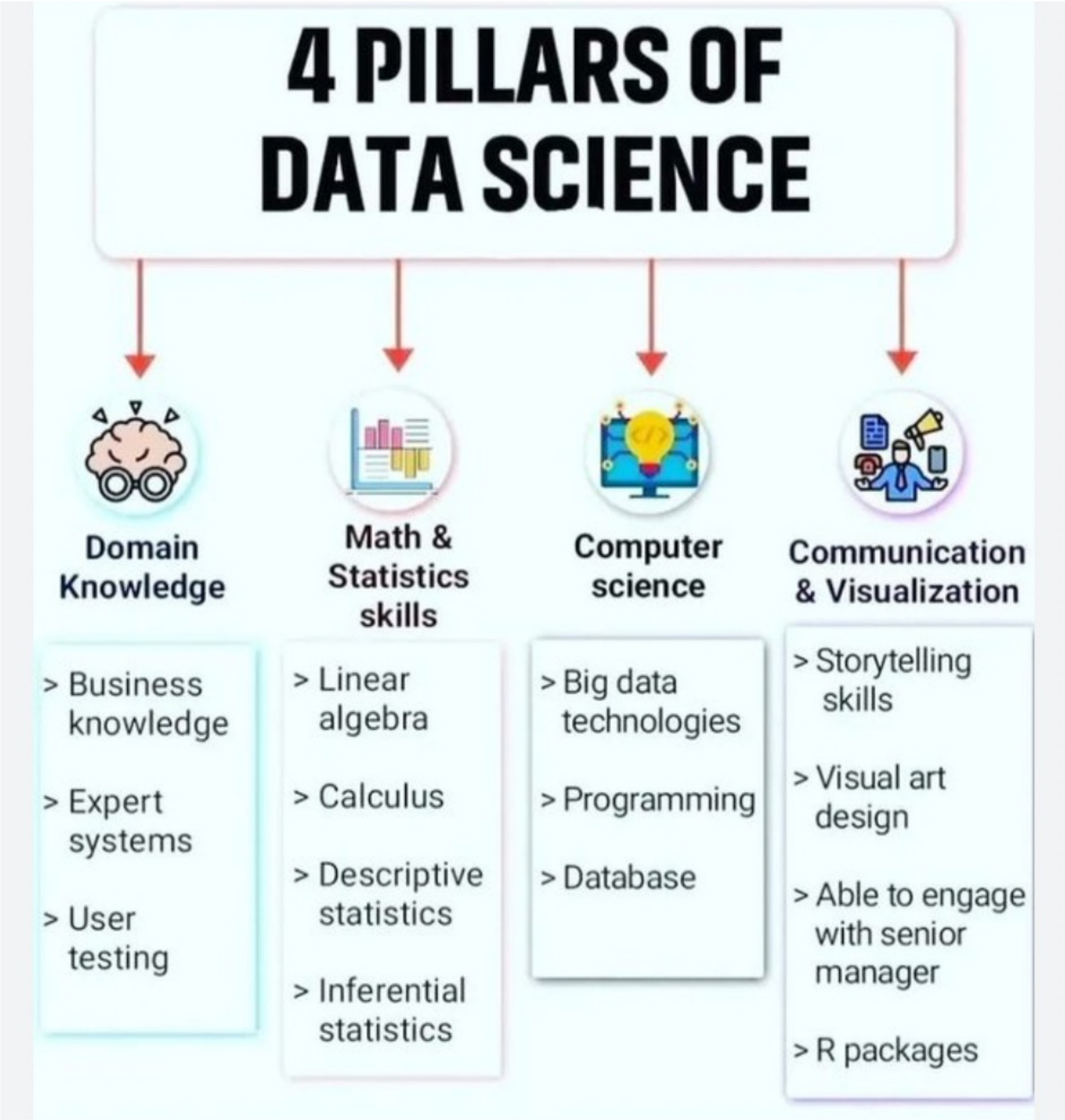
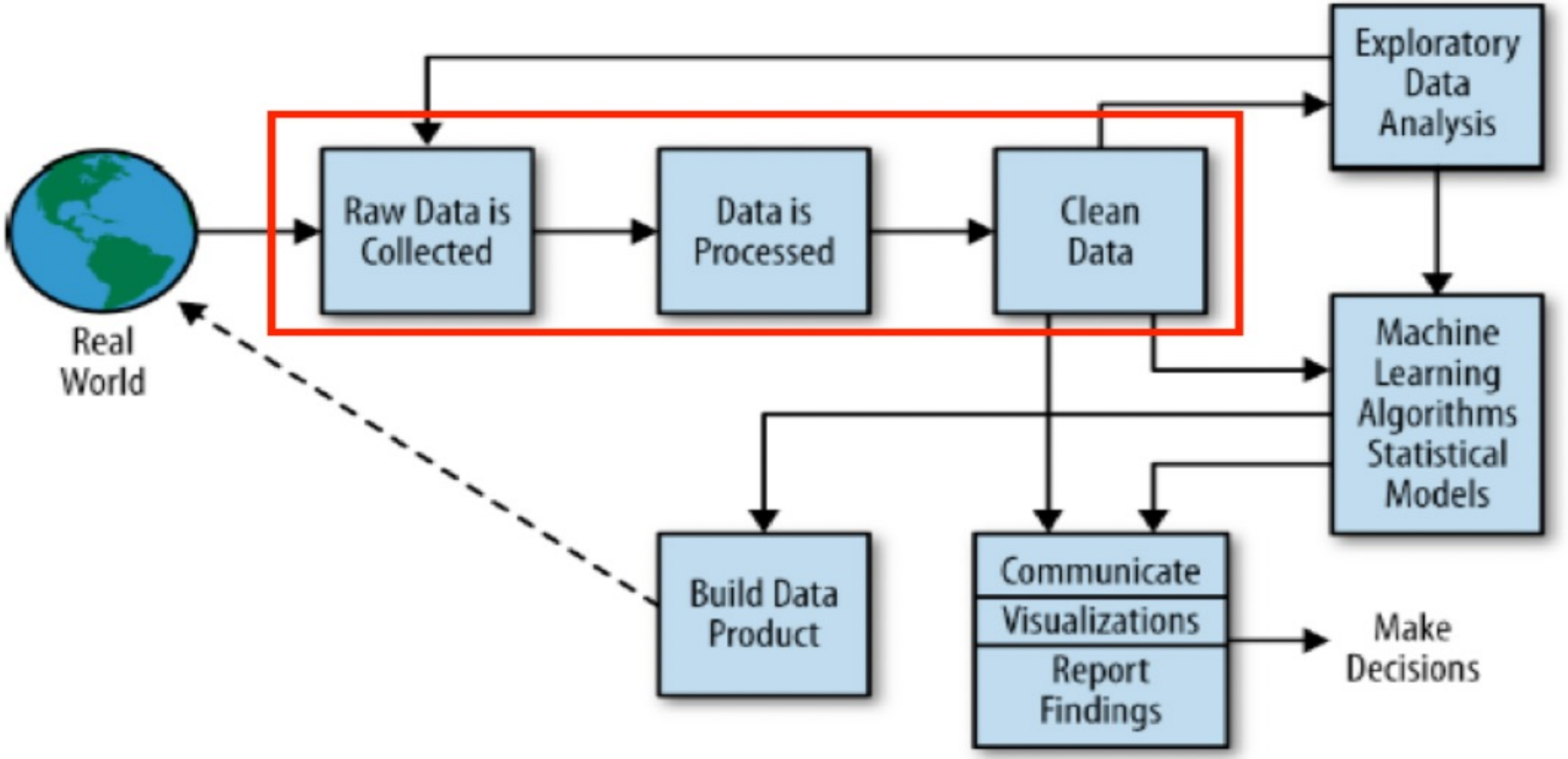
Week 1: Calculus in Data Science

Rabail Tahir



https://www.kaggle.com/discussions/getting-started/352351

The Data Science Process



Source: Doing Data Science
Rachel Schutt & Cathy O'Neil

Recommended Books

Applied Calculus 5th Edition:

- https://learning.oreilly.com/library/view/applied-calculus-5th/9781118174920/

Essential Math for Data Science:

- https://learning.oreilly.com/library/view/essential-math-for/9781098102920/

Mathematics in Data Science

- Calculus
- Statistics
- Probability
- Linear Algebra

What is Calculus?

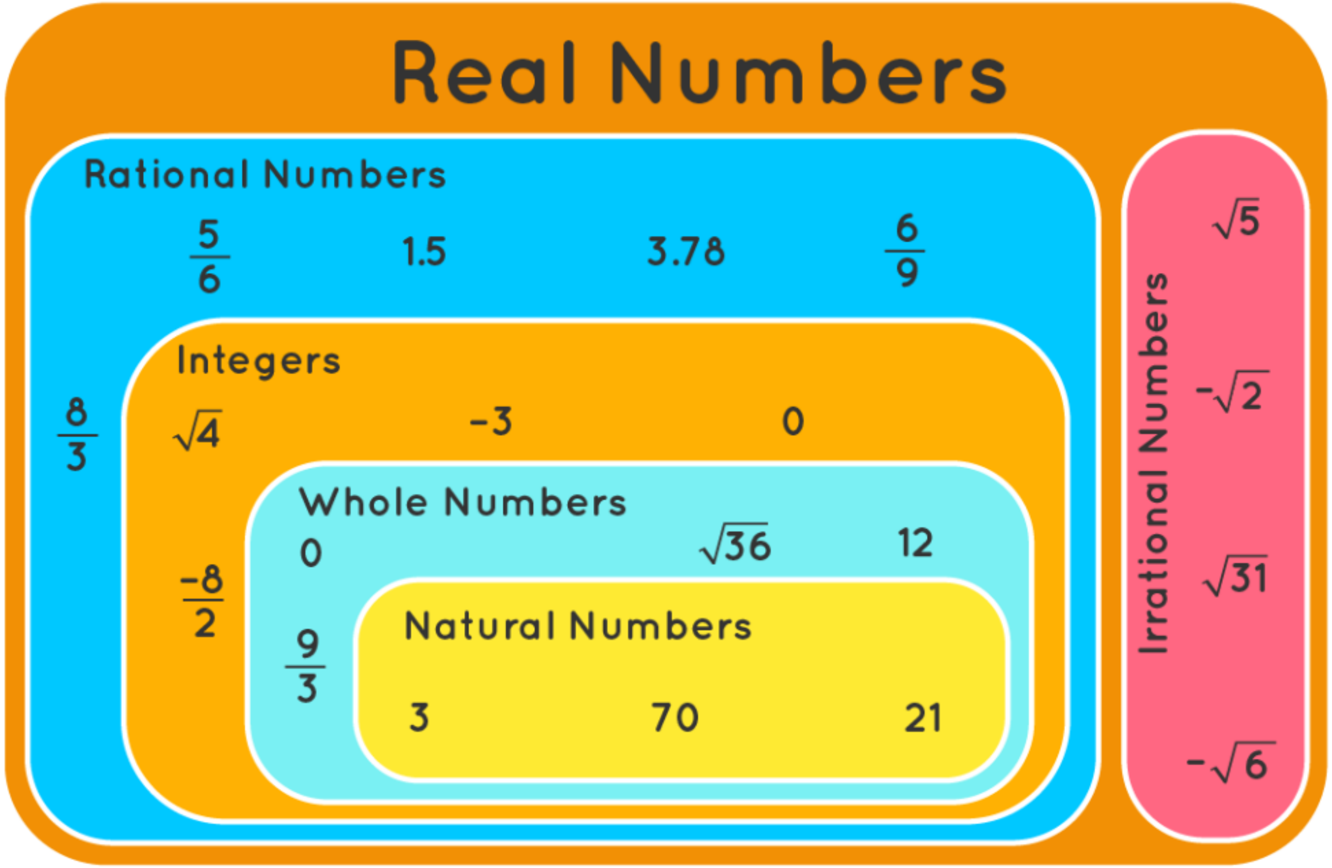
- "Calculus is a branch of mathematics that involves the study of rates of change"
- It is derived from the word "Calculi" , which means pebbles.
- It is the tracking of small changes.

Application of Calculus in Data Science

- Some of the concepts in Calculus are used to study changes and are applied in Machine learning algorithms.
- Calculus concepts to be covered in this lecture:
  - Number theory
  - Equations, Functions
  - Logarithm, exponential, polynomial functions, rational numbers.
  - Basic geometry and theorems, trigonometric identities.
  - Real and complex numbers and basic properties.
  - Series, sums, and inequalities.
  - Graphing and plotting, Cartesian and polar co-ordinate systems, conic sections.
  - Differentiation/ Integration

Number Theory and Number Systems

- Number Theory is the study of Number systems. We need Number systems we understand the numbers around us.
- Number systems summary:
  - Natural Numbers (1,2,3,4,5...) [Only positive numbers are included]
  - Integers (…-2,-1,0,1,2...) [Positive and Negative natural numbers including 0]
  - Rational Numbers (p/q where q is not equal to 0, any number that can be expressed as a fraction
  - Irrational Numbers
  - Irrational numbers cannot be expressed as a fraction. This includes the famous π, square roots of certain numbers.


Real Numbers
Rational Numbers
$\frac{5}{6}$
1.5
3.78
$\frac{6}{9}$
$\frac{8}{3}$
Integers
$\sqrt{4}$
-3
0
$\frac{-8}{2}$
Whole Numbers
0
$\sqrt{36}$
12
$\frac{9}{3}$
Natural Numbers
3
70
21
Irrational Numbers
$\sqrt{5}$
$-\sqrt{2}$
$\sqrt{31}$
$-\sqrt{6}$

Order of operations

- Ordering parentheses, exponents, multiplication, division, addition, and subtraction.
- For example consider this expression:

$$2 \times \frac{(3+2)^2}{5} - 4$$

Order of Operations

First we evaluate the parentheses (3 + 2), which equals 5:

$$2 \times \frac{(\mathbf{5})^2}{5} - 4$$

Next we solve the exponent, which we can see is squaring that 5 we just summed. That is 25:

$$2 \times \frac{\mathbf{25}}{5} - 4$$

Next up we have multiplication and division. The ordering of these two is swappable since division is also multiplication (using fractions). Let's go ahead and multiply the 2 with the $\frac{25}{5}$, yielding $\frac{50}{5}$:

$$\frac{50}{5} - 4$$

Order of Operations contd

Next we will perform the division, dividing 50 by 5, which will yield 10:

$$\mathbf{10} - 4$$

And finally, we perform any addition and subtraction. Of course, $10 - 4$ is going to give us 6:

$$10 - 4 = 6$$

Functions

- Variables are named values/quantities
- Function is defined as the relationship between 2 or more variables.

Take this simple linear function:

$$y = 2x + 1$$

For any given x-value, we solve the expression with that x to find y.

Functions

Table 1-1. Different values for $y = 2x + 1$

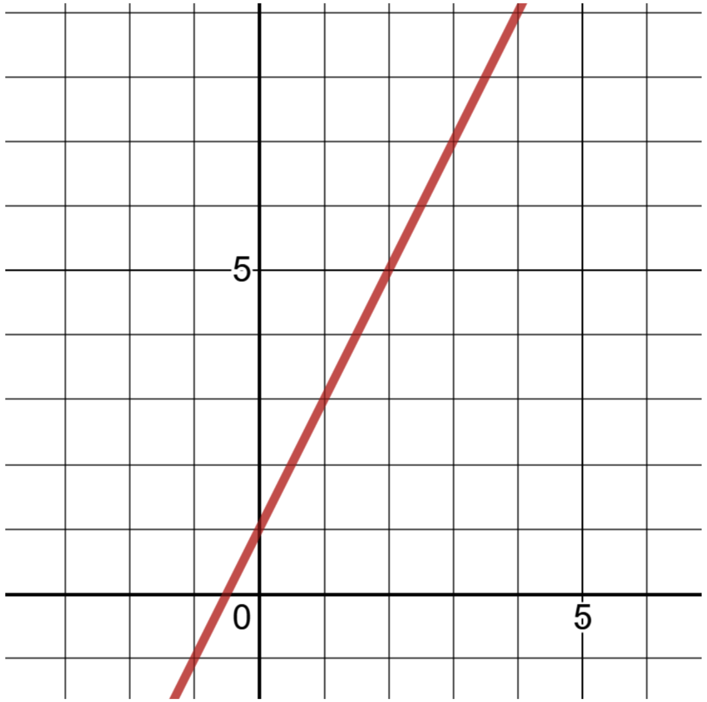
x	2x + 1	y
0	$2(0) + 1$	1
1	$2(1) + 1$	3
2	$2(2) + 1$	5
3	$2(3) + 1$	7

What are Functions used for in Data Science?

- Functions are useful because they model a predictable relationship between variables, such as how many fires *y* can we expect at *x* temperature.

Visualizing the Function

- Visualizing the function means to plot on a two-dimensional plane with two number lines (one for each variable) it is known as a *Cartesian plane*, *x-y plane*, or *coordinate plane*.



Summations

- A summation is expressed as a sigma Σ and adds elements together.
- For example:

$$\sum_{i=1}^{5} 2i = (2)1 + (2)2 + (2)3 + (2)4 + (2)5 = 30$$

Python code for summation:

```
x = [1, 4, 6, 2]
n = len(x)

summation = sum(10*x[i] for i in range(0,n))
print(summation)
```

Exponents

- *Exponents* multiply a number by itself a specified number of times.
- The *base* is the variable or value we are exponentiating, and the *exponent* is the number of times we multiply the base value. For the expression 23, 2 is the base and 3 is the exponent.

Rules of Exponent

Exponent Rules For $a \neq 0, b \neq 0$	
Product Rule	$a^x \times a^y = a^{x+y}$
Quotient Rule	$a^x \div a^y = a^{x-y}$
Power Rule	$\left(a^x\right)^y = a^{xy}$
Power of a Product Rule	$(ab)^x = a^x b^x$
Power of a Fraction Rule	$\left(\frac{a}{b}\right)^x = \frac{a^x}{b^x}$
Zero Exponent	$a^0 = 1$
Negative Exponent	$a^{-x} = \frac{1}{a^x}$
Fractional Exponent	$a^{\frac{x}{y}} = \sqrt[y]{a^x}$

Logarithms

- A *logarithm* is a math function that finds a power for a specific number and base.

Start your thinking by asking "2 raised to *what power* gives me 8?" One way to express this mathematically is to use an *x* for the exponent:

$$2^x = 8$$

We intuitively know the answer, $x = 3$, but we need a more elegant way to express this common math operation. This is what the $log()$ function is for.

$$log_2 8 = x$$

As you can see in the preceding logarithm expression, we have a base 2 and are finding a power to give us 8. More generally, we can reexpress a variable exponent as a logarithm:

$$a^x = b$$

$$log_a b = x$$

Properties for exponents and logarithms

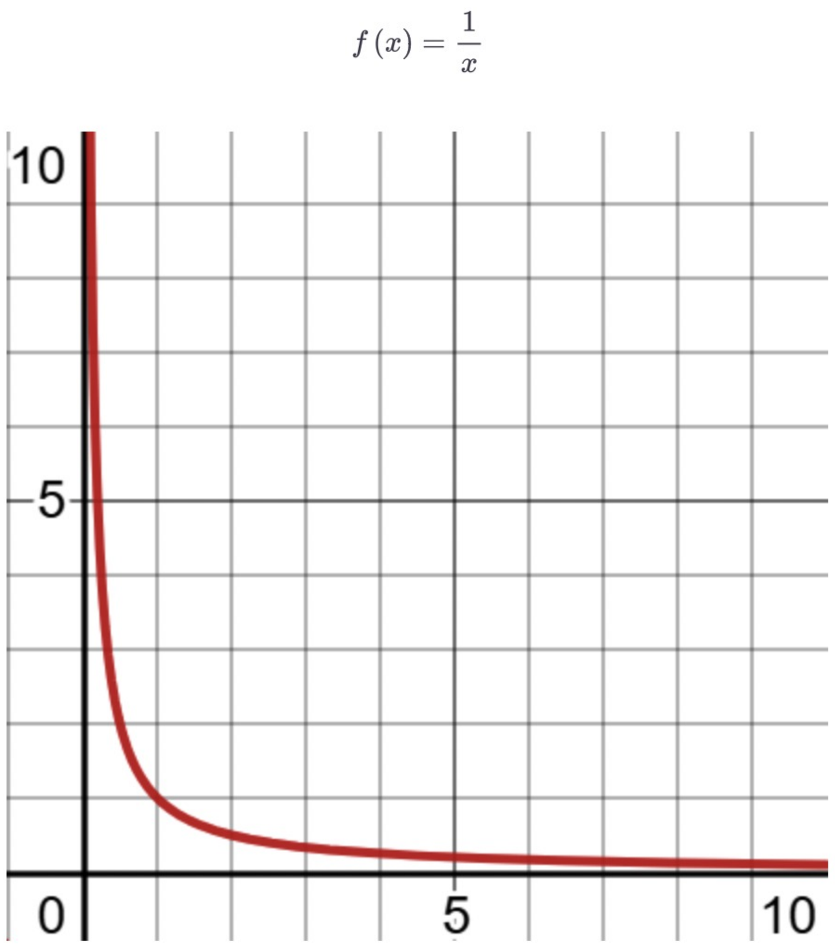
Operator	Exponent property	Logarithm property
Multiplication	$x^m \times x^n = x^{m+n}$	$log(a \times b) = log(a) + log(b)$
Division	$\frac{x^m}{x^n} = x^{m-n}$	$log\left(\frac{a}{b}\right) = log\,(a) - log\,(b)$
Exponentiation	$(x^m)^n = x^{mn}$	$log\,(a^n) = n \times log\,(a)$
Zero Exponent	$x^0 = 1$	$log(1) = 0$
Inverse	$x^{-1} = \frac{1}{x}$	$log\left(x^{-1}\right) = log\left(\frac{1}{x}\right) = -log\,(x)$

Euler's number and natural logarithms

- Euler's number is a special number just like pie and has a value of 2.71828
- A property of Euler’s number is its exponential function is a derivative to itself, which is convenient for exponential and logarithmic functions.
- When we use e as our base for a logarithm, we call it a *natural logarithm*. Depending on the platform, we may use ln() instead of log() to specify a natural logarithm. So rather than express a natural logarithm expressed as loge10 to find the power raised on e to get 10, we would shorthand it as ln(10).

Limits

- "Limits in maths are defined as the values that a function approaches the output for the given input values"



Laws of Limits

Assume that $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ exist and that c is any constant. Then,

1. $\lim_{x \to a} c = c$
2. $\lim_{x \to a} x = a$
3. $\lim_{x \to a} [c f(x)] = c \lim_{x \to a} f(x)$
4. $\lim_{x \to a} [f(x) \pm g(x)] = \lim_{x \to a} f(x) \pm \lim_{x \to a} g(x)$
5. $\lim_{x \to a} [f(x) g(x)] = \lim_{x \to a} f(x) \cdot \lim_{x \to a} g(x)$
6. $\lim_{x \to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}, \quad \lim_{x \to a} g(x) \neq 0$
7. $\lim_{x \to a} [f(x)]^n = [\lim_{x \to a} f(x)]^n$, where $n \in \mathrm{N}$
8. $\lim_{x \to a} x^n = a^n$
9. $\lim_{x \to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x \to a} f(x)}$, $\lim_{x \to a} f(x) > 0$ if n is even.
10. $\lim_{x \to a} [\ln f(x)] = \ln[\lim_{x \to a} f(x)]$, $\lim_{x \to a} f(x) > 0$

L'Hopital's Rule

If $\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{0}{0}$ or $\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\infty}{\infty}$, then $\lim_{x \to a} \frac{f'(x)}{g'(x)}$

Example

Find $\lim_{x\to -2} \dfrac{x+2}{x^2+3x+2}$

(step 1) **Plug in to evaluate the limit, if possible**

$$\lim_{x\to -2} \frac{x+2}{x^2+3x+2} = \frac{(-2)+2}{(-2)^2+3(-2)+2} = \frac{0}{0} \text{ (indeterminate)}$$

(step 2) **Apply L'Hospital's rule and reevaluate the limit**

$$\lim_{x\to -2} \frac{\frac{d}{dx}[x+2]}{\frac{d}{dx}[x^2+3x+2]} = \lim_{x\to -2} \frac{1}{2x+3}$$

$$\lim_{x\to -2} \frac{1}{2x+3} = \frac{1}{2(-2)+3} = \frac{1}{-1} = \boxed{-1}$$

Example

Find $\lim\limits_{x\to\infty}\dfrac{2x^2-5x+1}{3+x+6x^2}$

(step 1) **Plug in to evaluate the limit, if possible**

$$\lim_{x\to\infty}\frac{2(\infty)^2-5(\infty)+1}{3+(\infty)+6(\infty)^2}=\frac{\infty}{\infty}\quad\text{(indeterminate)}$$

(step 2) **Apply L'Hospital's rule and reevaluate the limit**

$$\lim_{x\to\infty}\frac{\frac{d}{dx}\left[2x^2-5x+1\right]}{\frac{d}{dx}\left[3+x+6x^2\right]}=\lim_{x\to-2}\frac{4x-5}{1+12x}$$

$$\lim_{x\to\infty}\frac{4x-5}{1+12x}=\frac{4(\infty)-5}{1+12(\infty)}=\frac{\infty}{\infty}\quad\text{(still indeterminate)}$$

(step 3) **Apply L'Hospital's rule again, and reevaluate the limit**

$$\lim_{x\to\infty}\frac{\frac{d}{dx}\left[4x-5x\right]}{\frac{d}{dx}\left[1+12x\right]}=\lim_{x\to\infty}\frac{4}{12}$$

$$\lim_{x\to\infty}\frac{4}{12}=\frac{4}{12}=\boxed{\frac{1}{3}}$$

Derivatives

- "A *derivative* tells the slope of a function, and it is useful to measure the rate of change at any point in a function."
- Derivatives are used in machine learning and various mathematical algorithms.
- A tangent is a line which touches a given function on any given point and tells us about the slope of a function on that particular point.
- We use this concept to cut the function into small pieces to study how it changes.

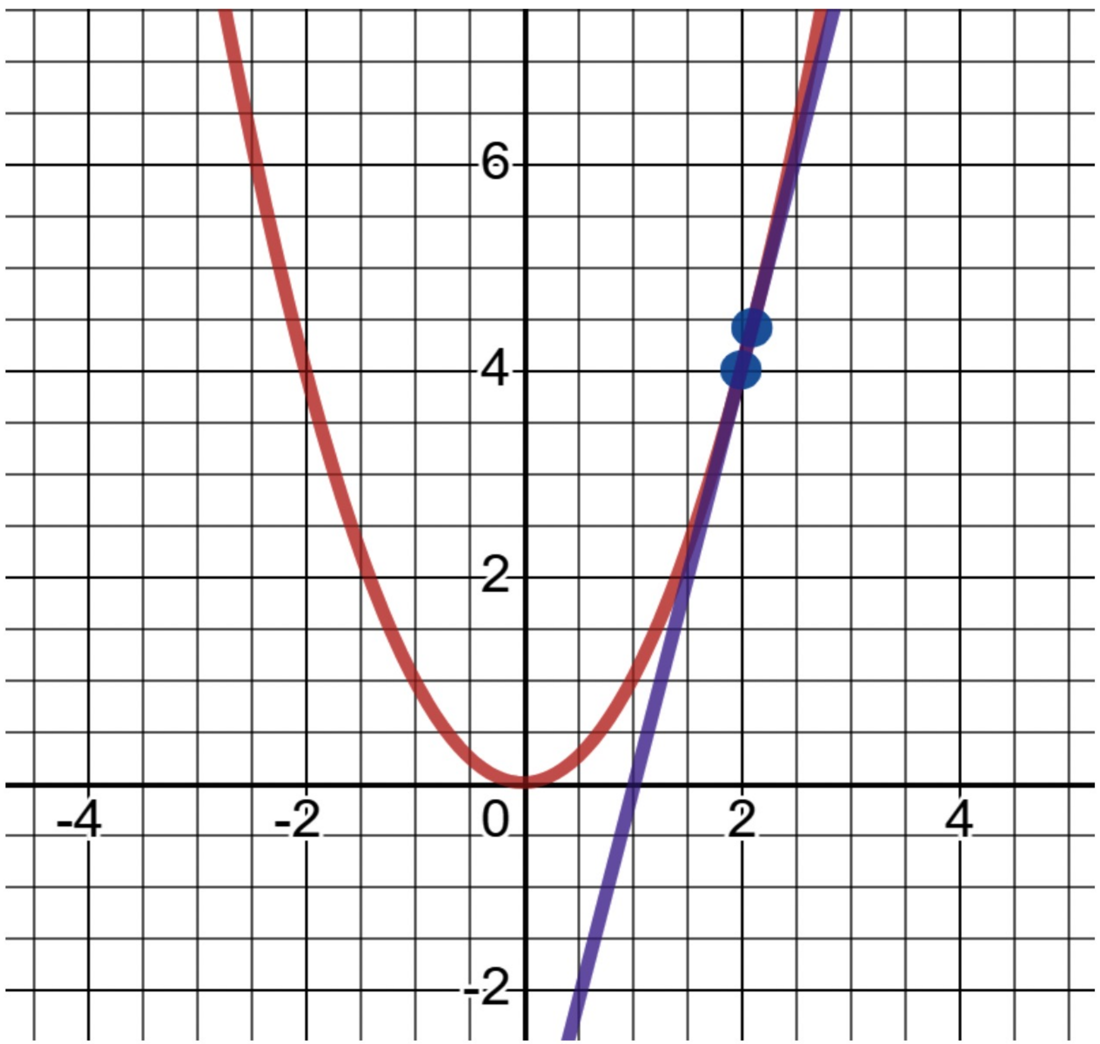
Calculating Slope

- Consider this graph for the function f(x)= x^2
- To calculate the slope m at the point of intersection

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$m = \frac{4.41 - 4.0}{2.1 - 2.0}$$

$$m = 4.1$$



The Derivative

1. Choose an interval

2. Find the raw change

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

4. Make your model perfect

3. Find the rate of change

Basic Derivatives Rules

Constant Rule: $\frac{d}{dx}(c) = 0$

Constant Multiple Rule: $\frac{d}{dx}[cf(x)] = cf'(x)$

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

Sum Rule: $\frac{d}{dx}[f(x) + g(x)] = f'(x) + g'(x)$

Difference Rule: $\frac{d}{dx}[f(x) - g(x)] = f'(x) - g'(x)$

Product Rule: $\frac{d}{dx}[f(x)g(x)] = f(x)g'(x) + g(x)f'(x)$

Quotient Rule: $\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$

Chain Rule: $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$

Partial Derivatives

- Partial Derivatives deal with functions that have multiple input variables.
- Rather than finding the slope on a one-dimensional function, we have slopes with respect to multiple variables in several directions
- And for each variable derivative, we assume that the others are being held constant.
- Let's look at a 3D graph of the function

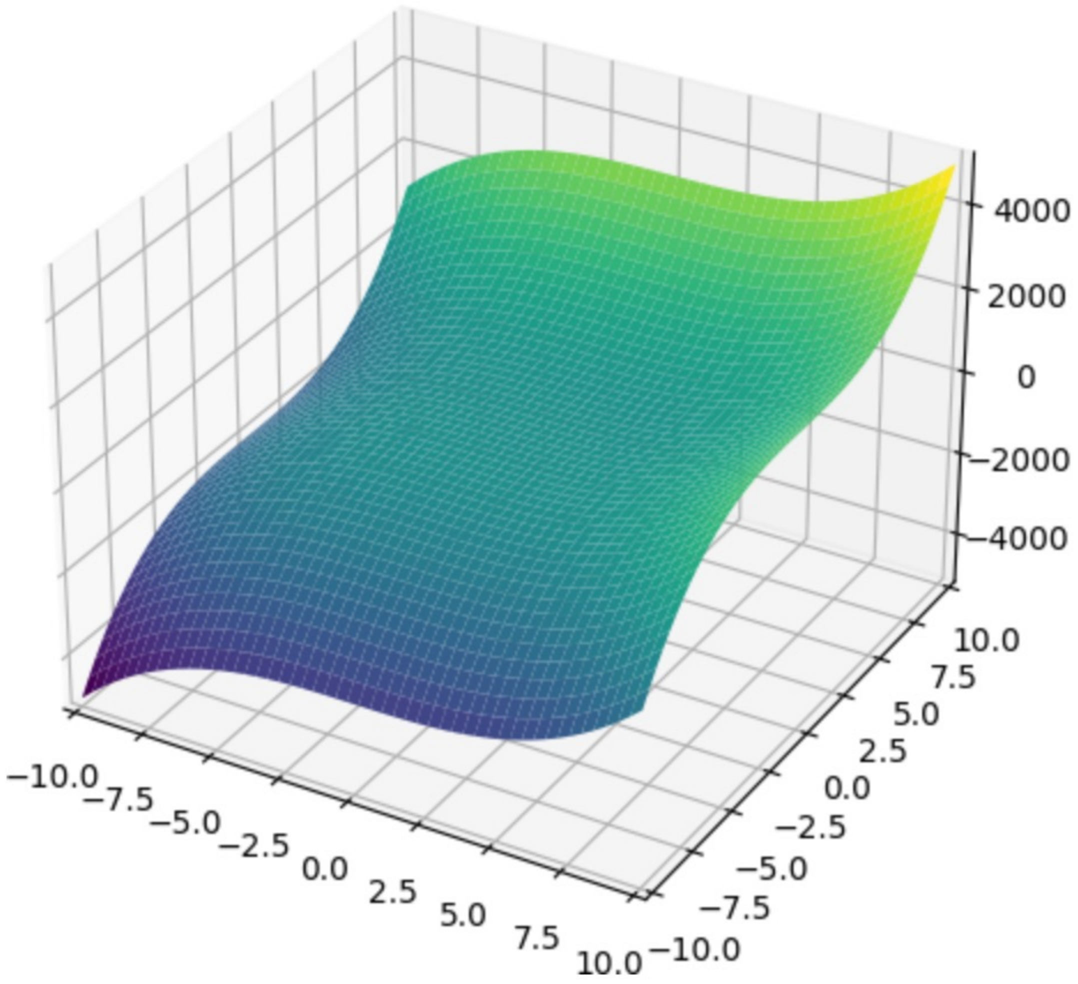
$$f(x, y) = 2x^3 + 3y^3$$

Calculating derivatives of x and y

$$f(x, y) = 2x^3 + 3y^3$$

$$\frac{d}{dx} 2x^3 + 3y^3 = 6x^2$$

$$\frac{d}{dy} 2x^3 + 3y^3 = 9y^2$$


4000
2000
0
−2000
−4000
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0

The Chain Rule

- The Chain rule is useful in building neural networks.
- Suppose you have 2 functions:

$$y = x^2 + 1$$

$$z = y^3 - 2$$

The Chain rule

Both functions are linked as the y variable is common between them.

We then substitute for y

$$z = \left(x^2 + 1\right)^3 - 2$$

So our derivative for z with respect to x is $6x\left(x^2 + 1\right)^2$:

$$\frac{dz}{dx}\left(\left(x^2 + 1\right)^3 - 2\right)$$

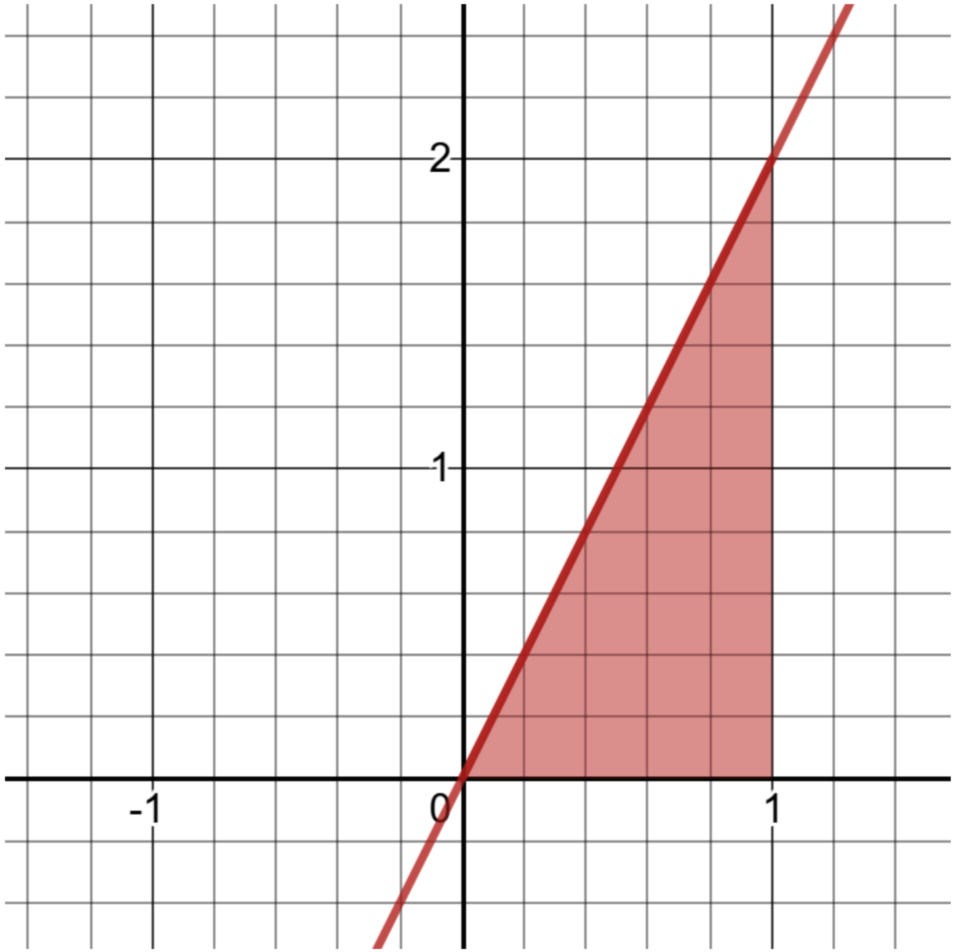
$$= 6x\left(x^2 + 1\right)^2$$

The Chain rule

$$\frac{dz}{dx} = \frac{dz}{dy} \times \frac{dy}{dx}$$

Integrals

- The opposite of derivative is an integral
- It is used to find the area under the curve for a given range
- For example, we have function $f(x) = 2x$ and we want to find the area between 0 and 1


2
1
-1
0
1

- If you recall basic geometry formulas, the area *A* for a triangle is

$$A = \frac{1}{2}bh$$

- where *b* is the length of the base and *h* is the height. We can visually spot that b=1 and h=2. So plugging into the formula, we get for our area 1.0.

What is the area between 0 and 1, as shaded in this case?

$$f(x) = x^2 + 1.$$

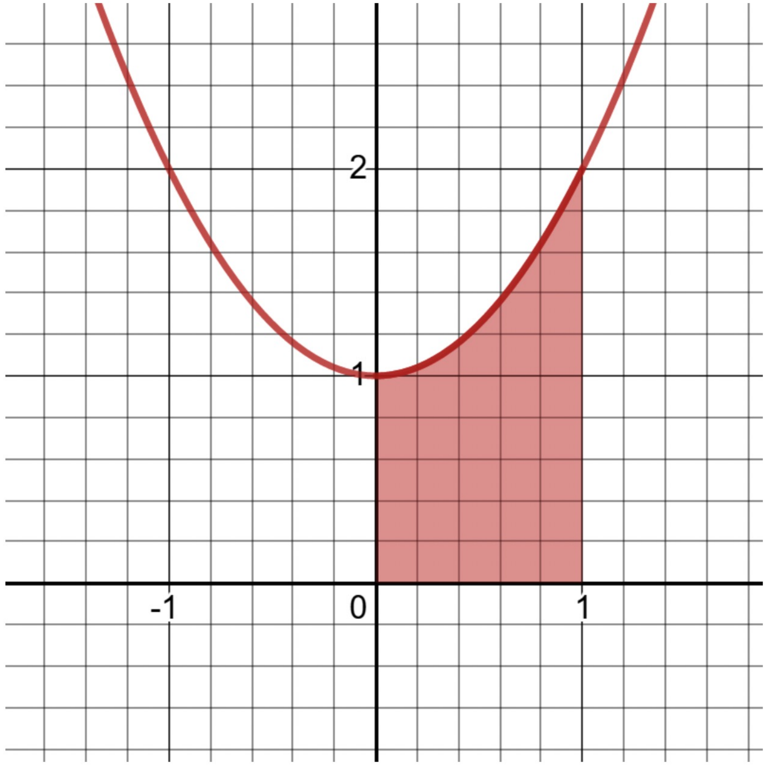
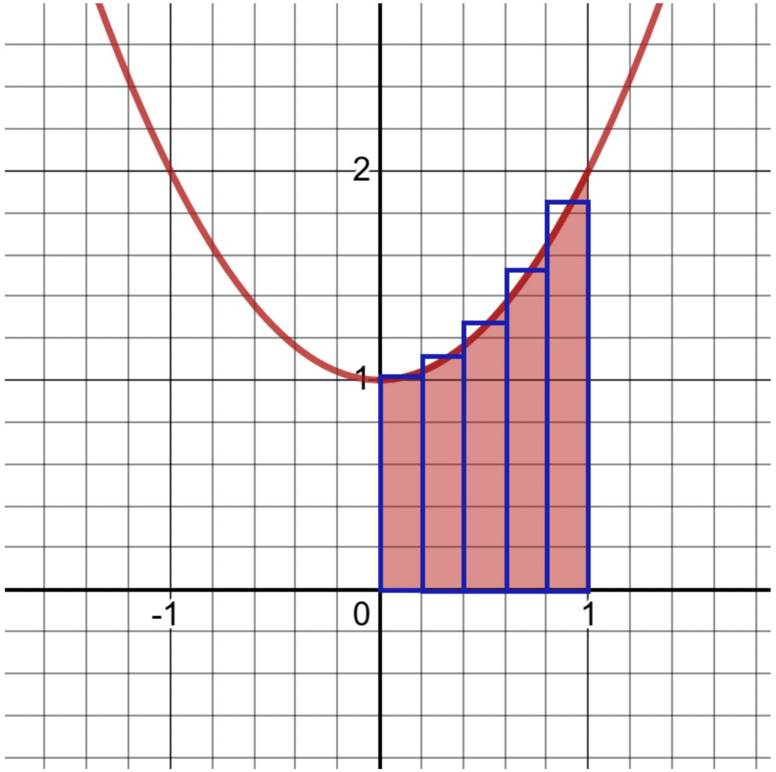


Figure 1.10. Calculating area under nonlinear functions is less straightforward

- The area of a rectangle is A=length×width, so we could easily sum the areas of the rectangles. Would that give us a good approximation of the area under the curve?



Any questions?