DS Workshop

Hyperiondev

Calculus for Data Science

Welcome

Your Lecturer for this session

Sanana Mwanawina

Workshop - Housekeeping

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment - please engage accordingly.
- No question is daft or silly - **ask them!**
- There are Q/A sessions midway and at the end of the session, should you wish to ask any follow-up questions.
- You can also submit questions here: www.hyperiondev.com/sbc4-ds-questions
- For all non-academic questions, please submit a query: www.hyperiondev.com/support
- Report a safeguarding incident: hyperiondev.com/safeguardreporting
- We would love your feedback on lectures and workshops: https://hyperionde.wufoo.com/forms/zsqv4m40ui4i0q/

GitHub repo

Go to: github.com/HyperionDevBootcamps

Then click on the “**C4_DS_lecture_examples**” repository, do view or download the code.

Objectives

1. Understand what derivatives and integrals are
2. Get an overview of how calculus is used in Data Science

Introduction to Calculus for Data Science

What is calculus? It is a branch of mathematics that deals with derivatives and integrals of functions.

Why do we need to know about calculus?

1. Calculus is the driving force behind the “learning” part of Machine Learning
2. In situations where we observe changes in data or variables, calculus can allow us to measure and model these changes
3. Probability and Statistics are fundamental to Data Science. Calculus provides the mathematical framework to model these probabilistic relationships, for example, probability distributions
4. Calculus is crucial for optimization algorithms, which are widely used in Data Science.

There are many more reasons that motivate the need for Calculus in Data Science. Its importance cannot be overstated.

Calculus concepts and notation

Function notation:

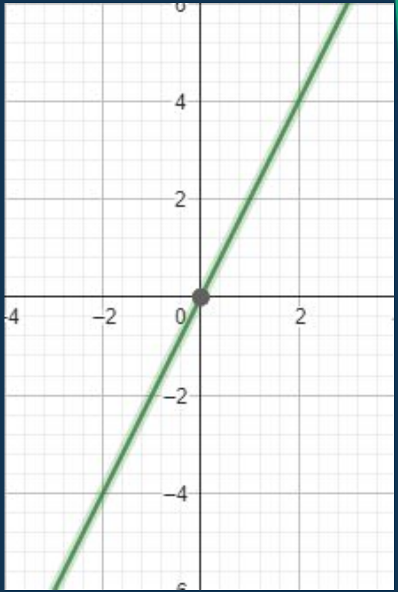
$$f(x) = 2x$$

This is a function f that takes an input x and doubles the input.

f(x) is also referred to as y.

$$y = f(x) = 2x$$

Example: when x = 2 we have that y = f(2) = 2(2) = 4



First-order derivatives

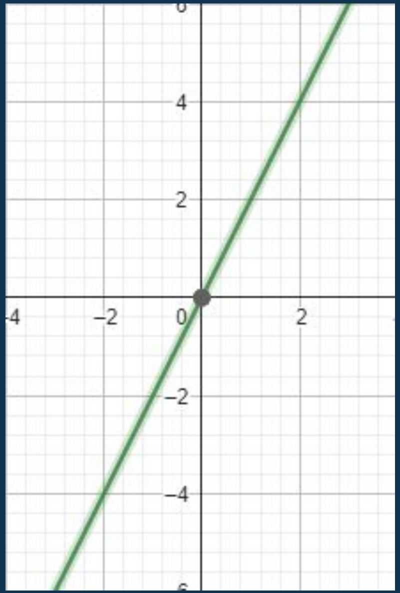
The derivative of a function gives us the instantaneous gradient of that function at the value of x. If we have a function f(x) and we take the derivative of it, we obtain the following function:

$$\frac{df}{dx}$$

which reads “the derivative of f with respect to x”. Let’s look at an example to understand this better.

First-order derivatives

Let f(x) = 2x. The graph on the right is f(x) plotted



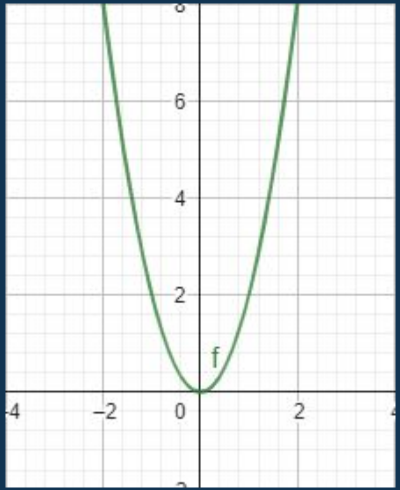
According to the definition of a derivative, we can take the derivative of f(x) and obtain a function, df/dx, that tells us the instant gradient of f at any point x.

In our case, df/dx = 2. Therefore, at any point x, the gradient of f will be 2.

f(x) is a straight line. So, using the general equation of a straight line y = mx + c, we know m = gradient = 2.

First-order derivatives

Let $f(x) = 2x^2$. The graph on the right is f(x) plotted



Finding the gradient is not as straightforward as it is with a straight line where we can "skip" finding the derivative and simply use the slope, denoted as m. However, by taking the derivative of the function f(x) and obtaining df/dx, we can deduce the gradient at different points along f(x).

In this case, $df/dx = 4x$.

So therefore, at the point $x = 1$, the gradient of f(x) is $4(1) = 4$. At the point $x = -2$, the gradient of f(x) is $4(-2) = -8$, and so on.

If you want to learn the differentiation rules, here is a helpful resource: [Derivative Rules (mathsisfun.com)]

First-order derivatives

Practically, how is this derivative helpful? In so many ways!

1. Gradient-based optimization such as gradient descent. More on this in a future workshop.

2. Model training. When training neural networks, we use derivatives in a process called backpropagation to improve the model's predictive accuracy.

   Useful resource:
   https://www.3blue1brown.com/lessons/neural-networks

Integration

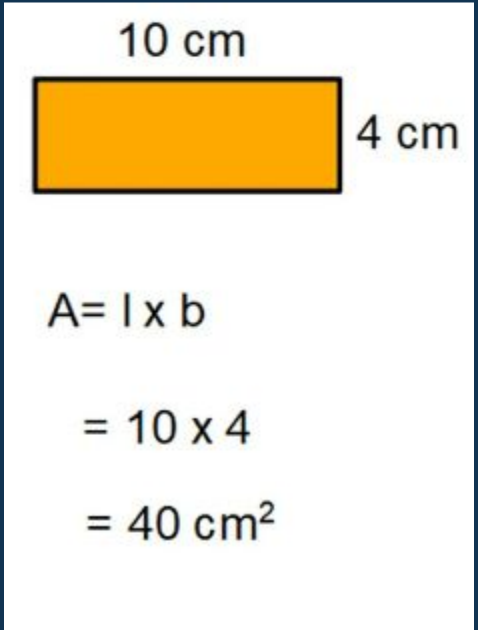
Another very useful Calculus technique is integration.

For the purposes of this workshop, we will focus on the application of integration when working with probability distributions. But before seeing the link between integration and probability distributions, let us see how we can use integration to find areas.

Finding the Area

Definition of area: a measure of a region's size on a surface. Depending on the shape of the region, we can calculate the area in different ways. For example, the area of a rectangle is:

$$A = \text{length} \times \text{breadth} = lb$$



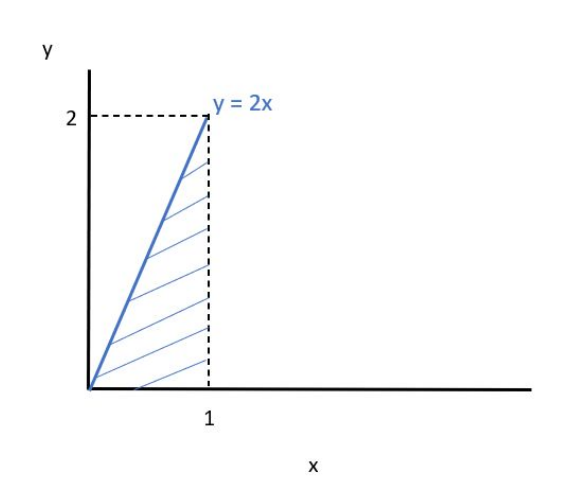
Using integration to find the area

Let's say we want to find the size of the area bounded by the line $y = 2x$, $x = 1$, and $y = 0$. If you look closely, the shape of the bounded area is a triangle. So, we are looking for the area of a triangle. You may be familiar with this formula:

$$\text{Area of a triangle} = \tfrac{1}{2}\,(\text{base})\,(\text{height})$$

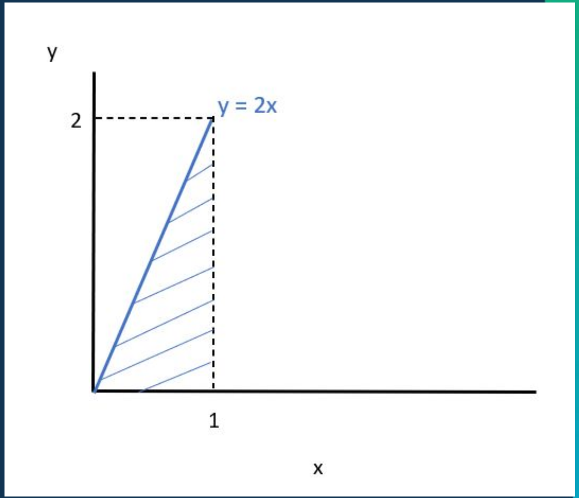
So, the area of the shaded region is $\tfrac{1}{2}(1)(2) = 1 \text{ unit}^2$

But we can use integration to find this area.



Using integration to find the area

The integral of our function, $y = f(x) = 2x$, is x^2.
A helpful resource for learning about how to calculate integrals:
[Introduction to Integration (mathsisfun.com)]



To get the area, we substitute the x values that bound our region of interest into our integral function, and then get the difference between the two.

Integral function is x^2
Lower bound x is 0
Upper bound x is 1

Area = $(1)^2 - (0)^2 = 1$ unit^2 , just as before when we used the formula

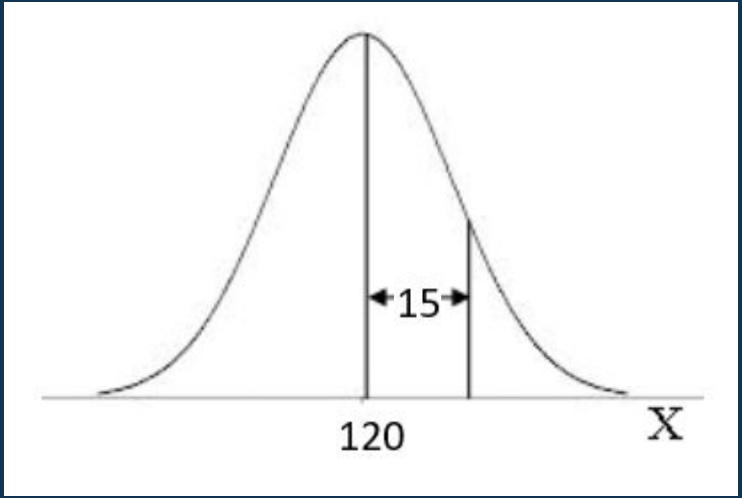
Integrals in Data Science

So, how is this “area under a line/curve” useful in Data Science? One application is when we want to find probabilities using probability density functions.

Think back to the normal distribution model that we had for the weights of females (Statistics workshop).

Normal Distribution

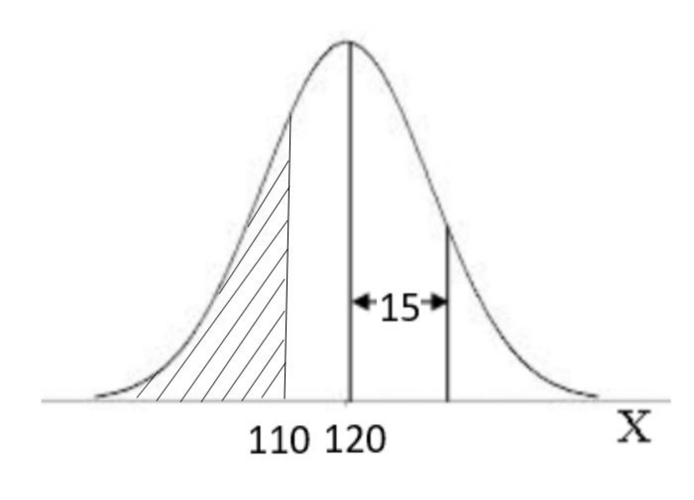
If X represents the weights of females, then $X \sim N(\mu = 120, \ \sigma^2 = 15^2)$



The normal distribution function is an example of a probability density function.

Area under the curve use in Data Science

Remember when we used Python to find Pr(X < 110)? In other words, the probability that a female weighs less than 110 lbs.



The cool thing about probability density functions is that the area below the curve is proportional to the probability associated with that range of x values.

So the area of the shaded region = 0.25 = Pr(X < 110)

Resource for diving into further Mathematics and Statistics

IntroSTAT by Les Underhill and Dave Bradfield:
https://open.uct.ac.za/bitstream/handle/11427/4150/INTROSTAT_ebook.pdf?sequence=1

Lecturer's advice:

- Before looking into probability distributions, make sure you cover the Random Variables section (Chapter 4)
- If you want to go through the resource chapter by chapter, start with Set Theory (chapter 2), then Probability Theory (chapter 3), then Exploring Data (chapter 1). Thereafter, you can follow the sequence of chapters as they appear
- IntroSTAT assumes the reader has a basic understanding of differentiation and integration

Hyperiondev

Q & A Section

Please use this time to ask any questions relating to the topic explained, should you have any

Hyperiondev

Thank you for joining us