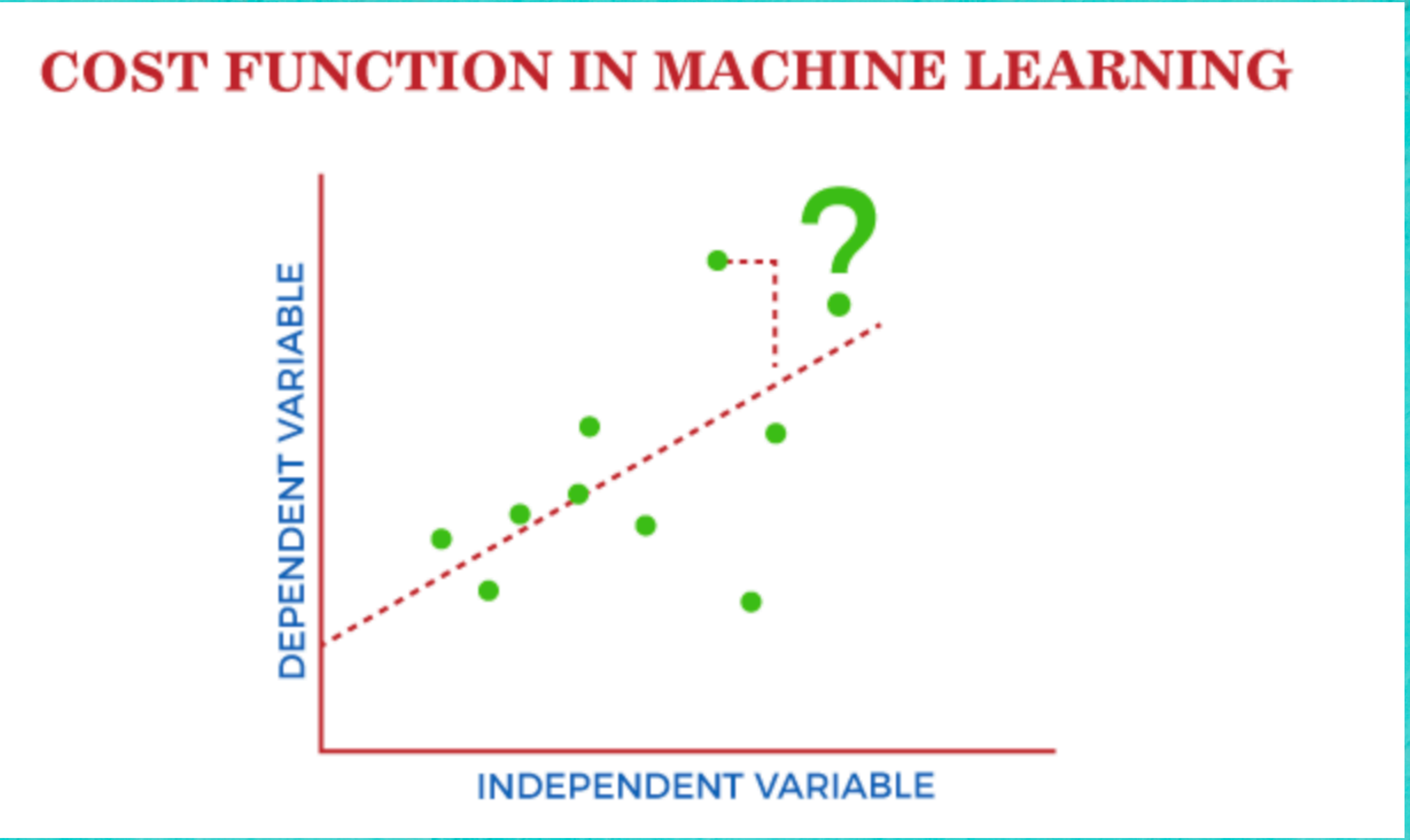
PRACTICAL EXERCISES
WEDNESDAY/THURSDAY
10TH AND 11TH /04–2024



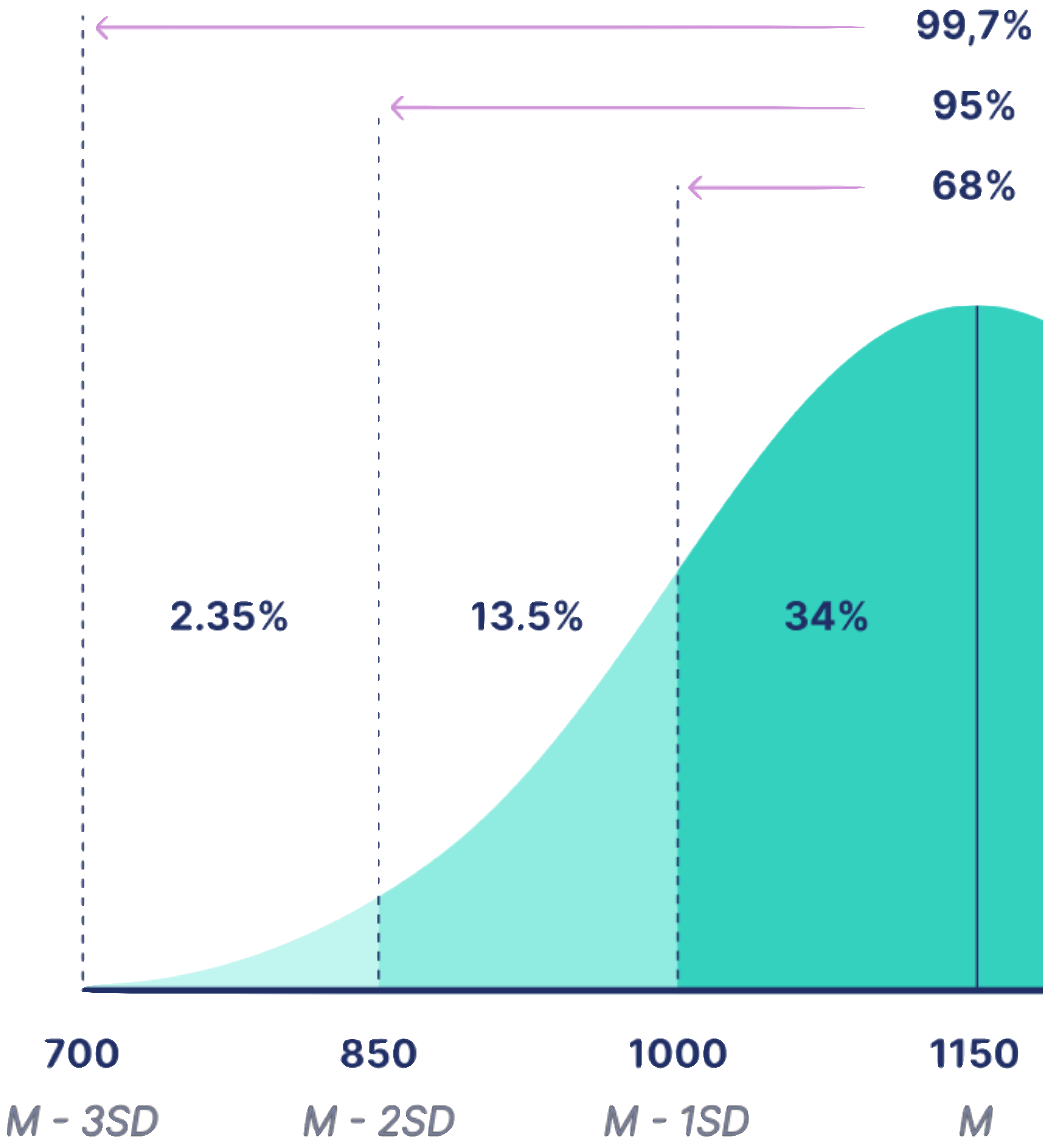
CLASS 8

UNDERSTANDING DERIVATIVES IN THE CONTEXT OF LINEAR REGRESSION
IS OLS = MACHINE LEARNING?

STUDENT INSTRUCTOR | PERNILLE HØJLUND BRAMS (AU650502@CAS.AU.DK)
COURSE | METHODS 2: THE GENERAL LINEAR MODEL

AGENDA

- First: A note on Matematik Tilvalg / Having math electives at CogSci
- Questions about the OLS problem
- Recapping
  - Residuals vs loss
  - Connections between OLS and broader frameworks (e.g. machine learning)
- Where do derivatives come in then??
- Exercises 📝
- Break
- More recap if time



https://www.scribbr.com/statistics/normal-distribution/

https://people.com/kevin-pietersen-says-seen-kate-middleton-last-couple-days-amid-conspiracy-theories-861123

4. semester	Cognitive Neuroscience (10 ECTS)	Methods 4: Bayesian Computational Modeling (10 ECTS)	Social and Cultural Dynamics (10 ECTS)
5. semester	Bachelor's project (15 ECTS)	TLV (15 ECTS)	
6. semester	TLV (30 ECTS)		

You decide what elective you want on 4th semester, so in a year :) and that's plenty time. If int. exchange: decide on 3rd semester

ORDINARY LEAST SQUARES (OLS)

ANY QUESTIONS ABOUT THE OLS RUN-THROUGH?

THE VIDEO: HTTPS://WWW.YOUTUBE.COM/WATCH?V=JXAY6JBSC0I

(YOU CAN ALWAYS EMAIL ME POST-CLASS IF YOUR QUESTIONS POP UP LATER OR IF YOU NEED FURTHER EXPLANATION:))

RESIDUALS AND LOSS

1 MINUTE ALONE —> 2 MINUTES IN GROUPS:

1) WHAT DO WE MEAN WHEN WE SAY ‘RESIDUAL’?
2) CAN WE USE THE RESIDUALS? IF SO, FOR WHAT?

RESIDUALS AND LOSS

DISCUSS FOR A MINUTE:

1) **WHAT DO WE MEAN WHEN WE SAY 'RESIDUAL'?**

DIFFERENCE BETWEEN OBSERVED VALUE OF Y AND VALUE PREDICTED BY THE REGRESSION MODEL

2) **CAN WE USE THE RESIDUALS? IF SO, FOR WHAT?**

DIAGNOSTIC TOOLS. METRIC FOR MODEL QUALITY. WAY TO FIND BEST MODEL IN MACHINE LEARNING (OPTIMISATION)

✨ DISCLAIMER ✨

WE’RE JUMPING A BIT AHEAD IN THE CURRICULUM NOW

(DON’T BE SCARED)

✨ GRADIENT DESCENT ✨

When we want to minimise the cost function (to find the best parameters = the best model), we do it via a process called **gradient descent.**

Gradient descent is a way to minimise the **cost function**.

✨ GRADIENT DESCENT ✨

When we want to minimise the cost function (to find the best parameters = the best model), we do it via a process called **gradient descent.**

Gradient descent is a way to minimise the **cost function**.

It involves taking the derivative of the **cost function** and using that to adjust the parameters of our neural network to reduce the cost.

✨ GRADIENT DESCENT ✨

When we want to minimise the cost function (to find the best parameters = the best model), we do it via a process called **gradient descent.**

Gradient descent is a way to minimise the **cost function**.

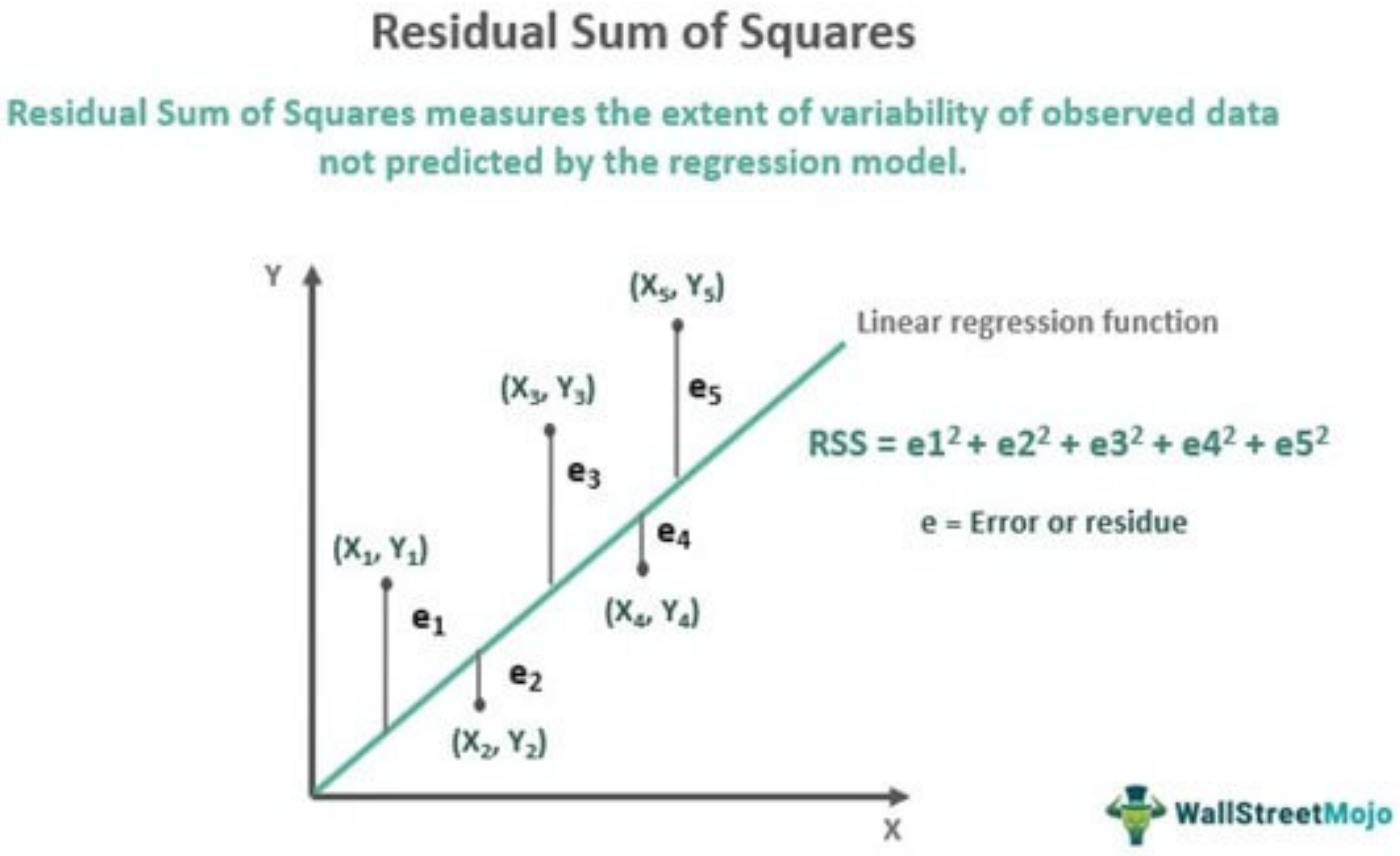
It involves taking the derivative of the **cost function** and using that to adjust the parameters of our neural network to reduce the cost.

Like in OLS where we want to minimise residuals, **gradient descent** is also about minimising residuals.

✨GRADIENT DESCENT✨MORE NOTES

In the context of linear regression, if you use gradient descent, **you're essentially trying to minimise the same cost function as OLS** (RSS or SSE) - but gradient descent allows for more flexibility.

This means gradient descent can be used not just for linear models but for **any** model where the cost function is differentiable with respect to its parameters.

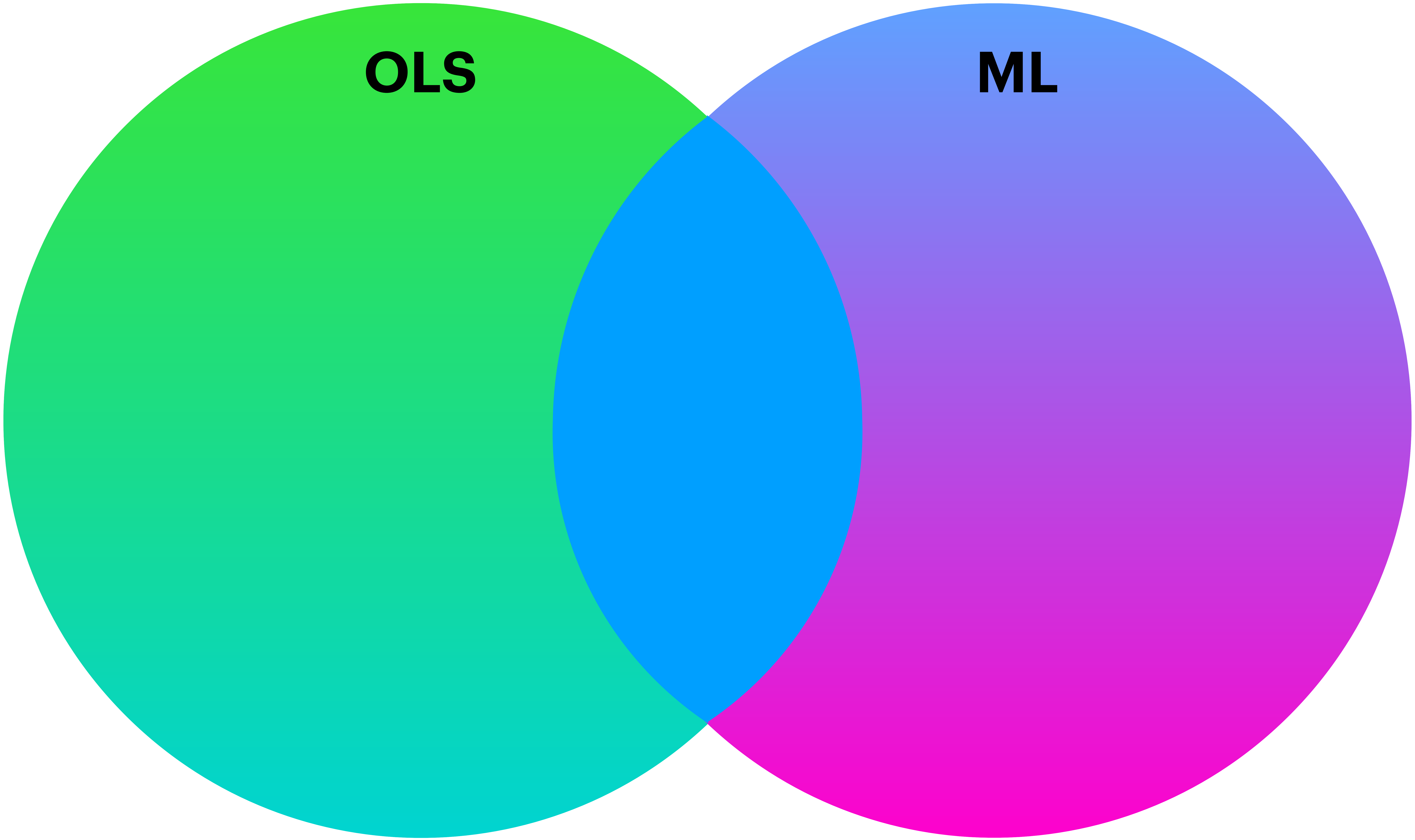


https://www.wallstreetmojo.com/residual-sum-of-squares/

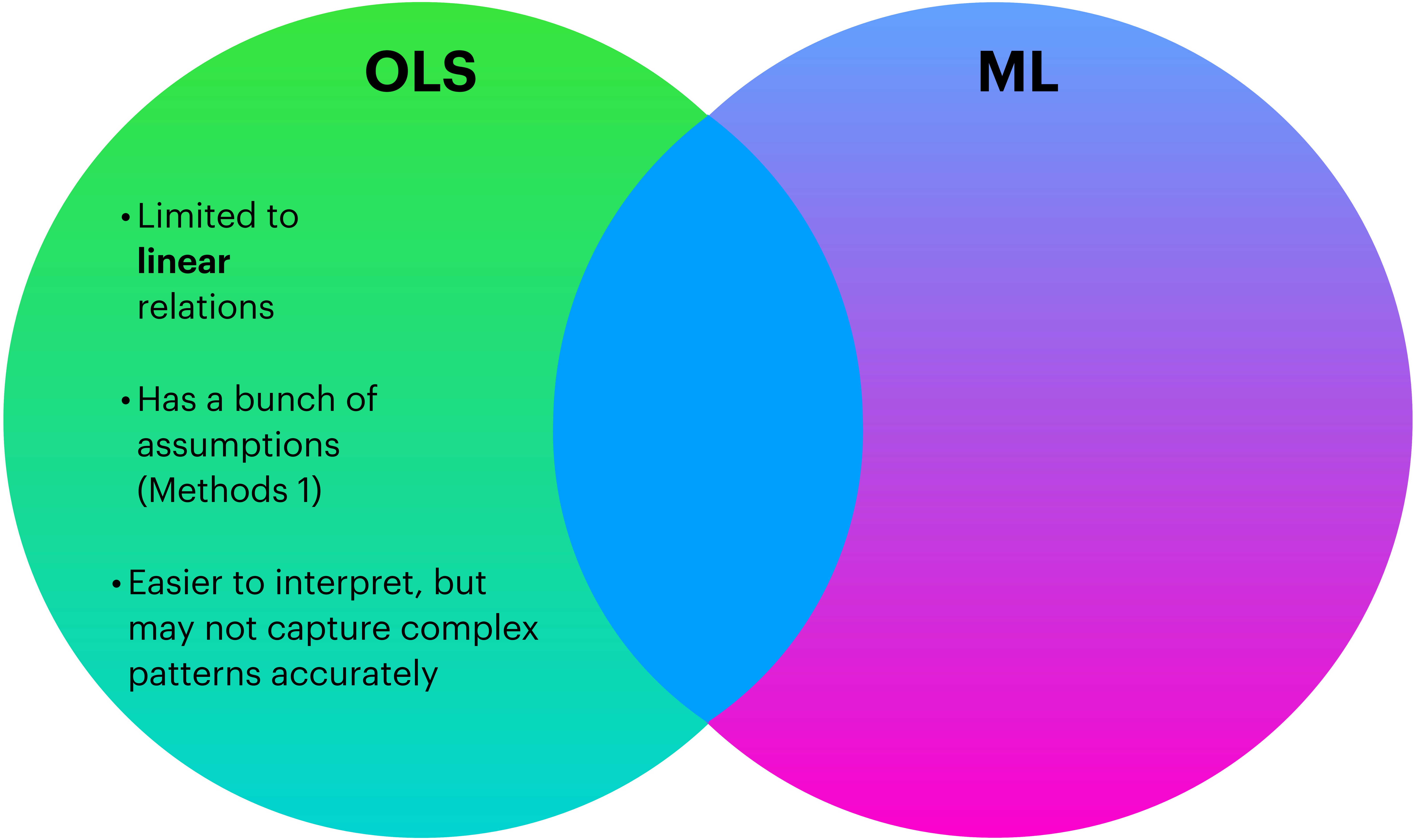
I'm an example of a cost function, and we actually use me in OLS ! :D

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

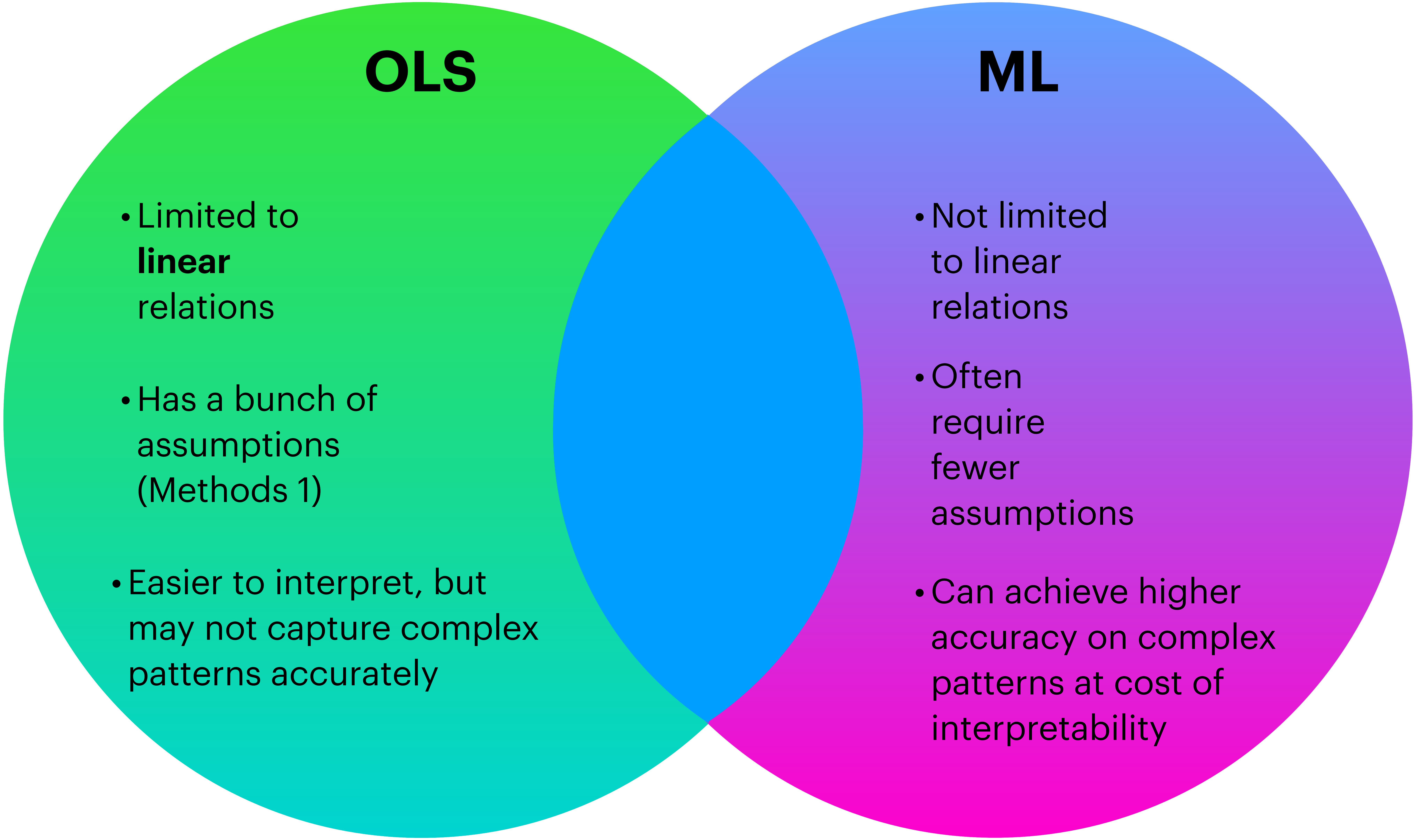
DO THE MECHANICS FROM OLS GENERALISE FURTHER?



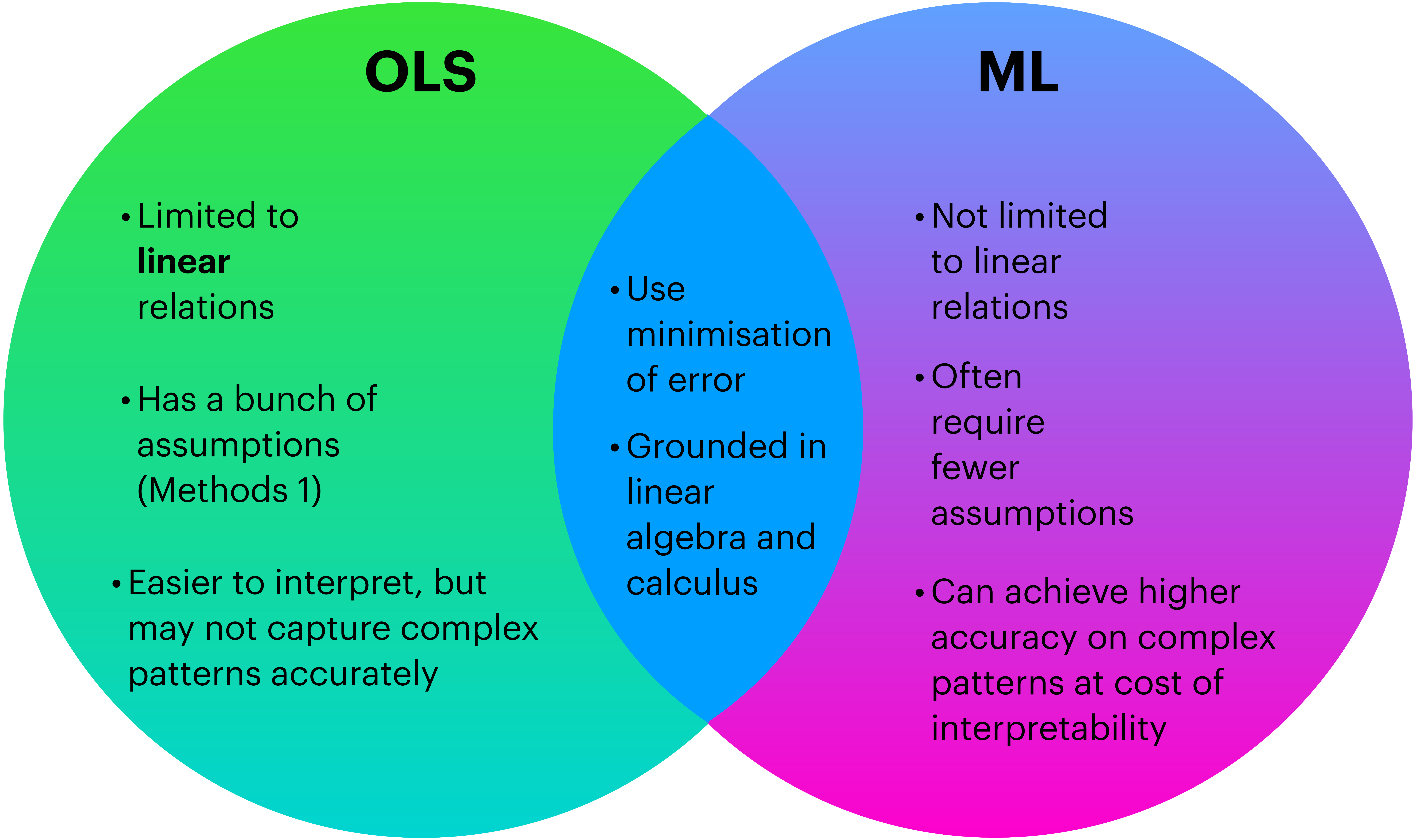
DO THE MECHANICS FROM OLS GENERALISE FURTHER?



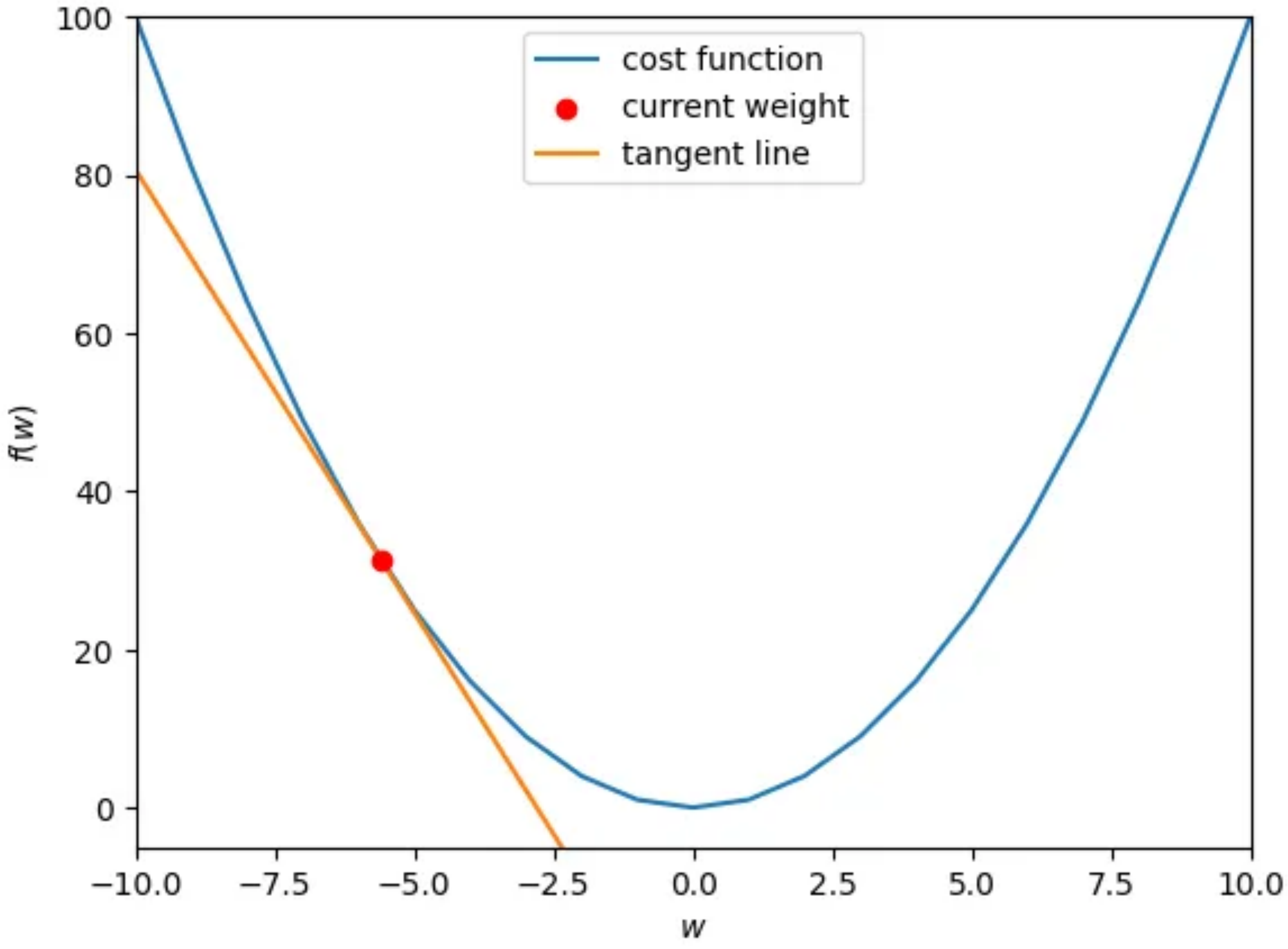
DO THE MECHANICS FROM OLS GENERALISE FURTHER?



DO THE MECHANICS FROM OLS GENERALISE FURTHER?



SO WHERE DO DERIVATIVES COME IN?

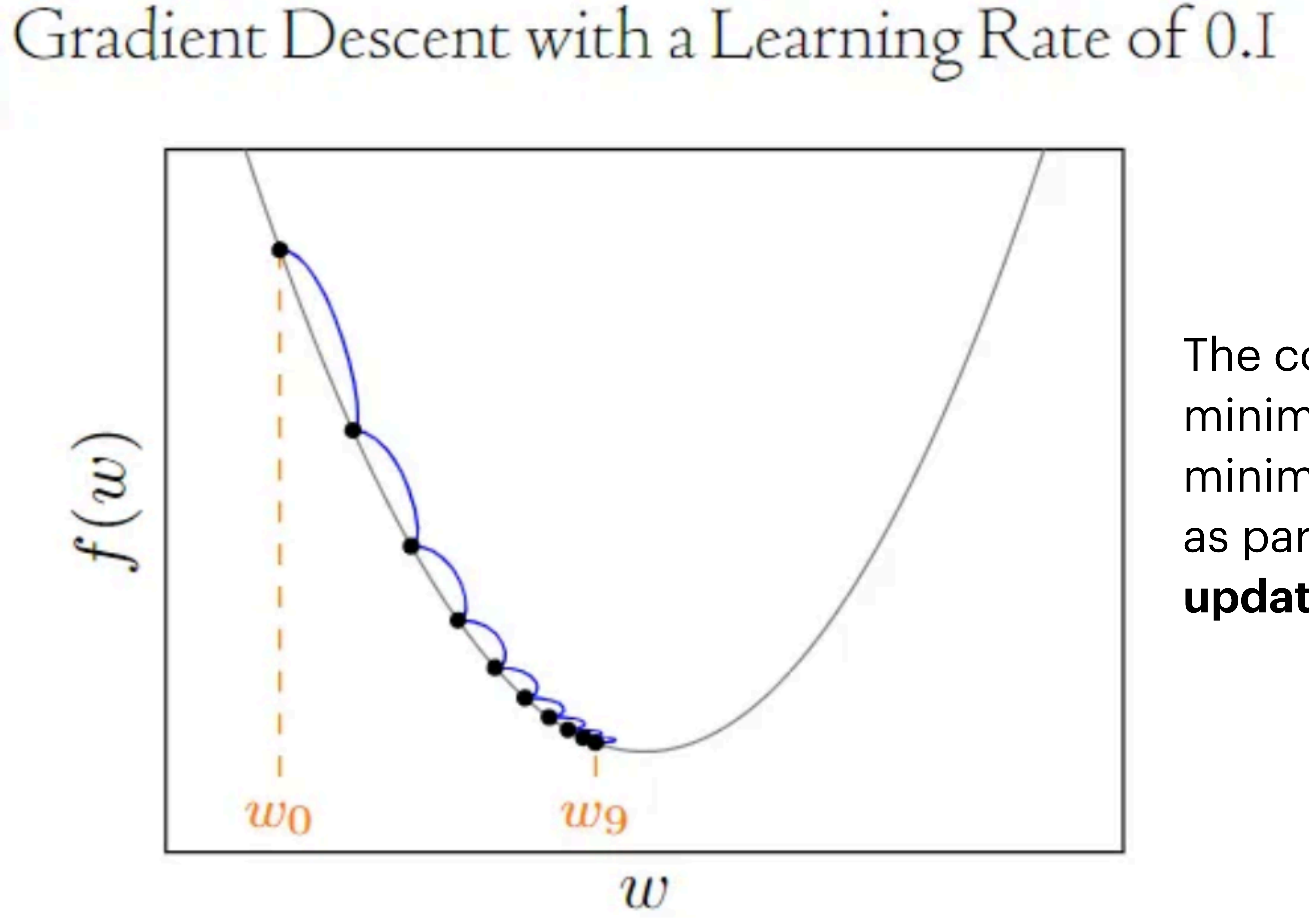


Derivative of each weight will reveal its direction and influence on the cost function.

Looking at the tangent here (the derivative of -11,2) - the current rate of change for the weight - this indicates the weight needs to move "downhill" or become more positive to reach a loss of 0.

https://medium.com/@hunter-j-phillips/a-simple-introduction-to-gradient-descent-1f32a08b0deb#:~:text=By%20using%20partial%20derivatives%2C%20a,influence%20on%20the%20cost%20function.

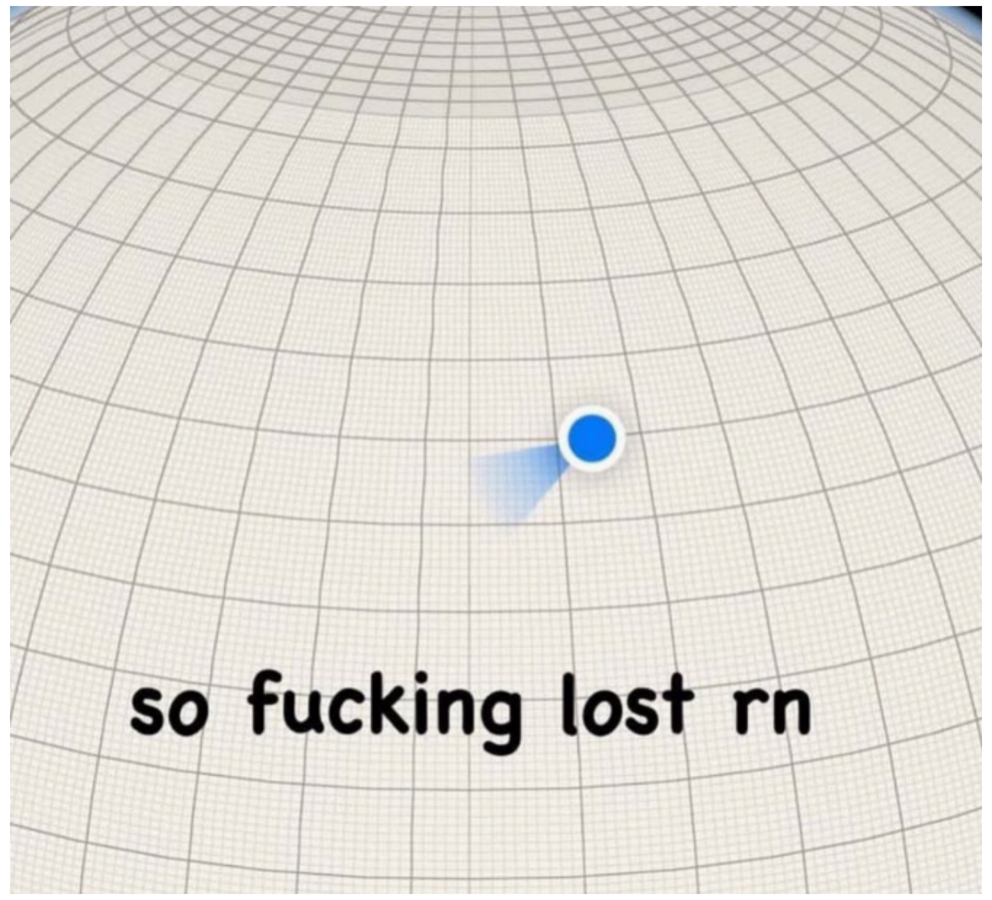
SO WHERE DO DERIVATIVES COME IN?



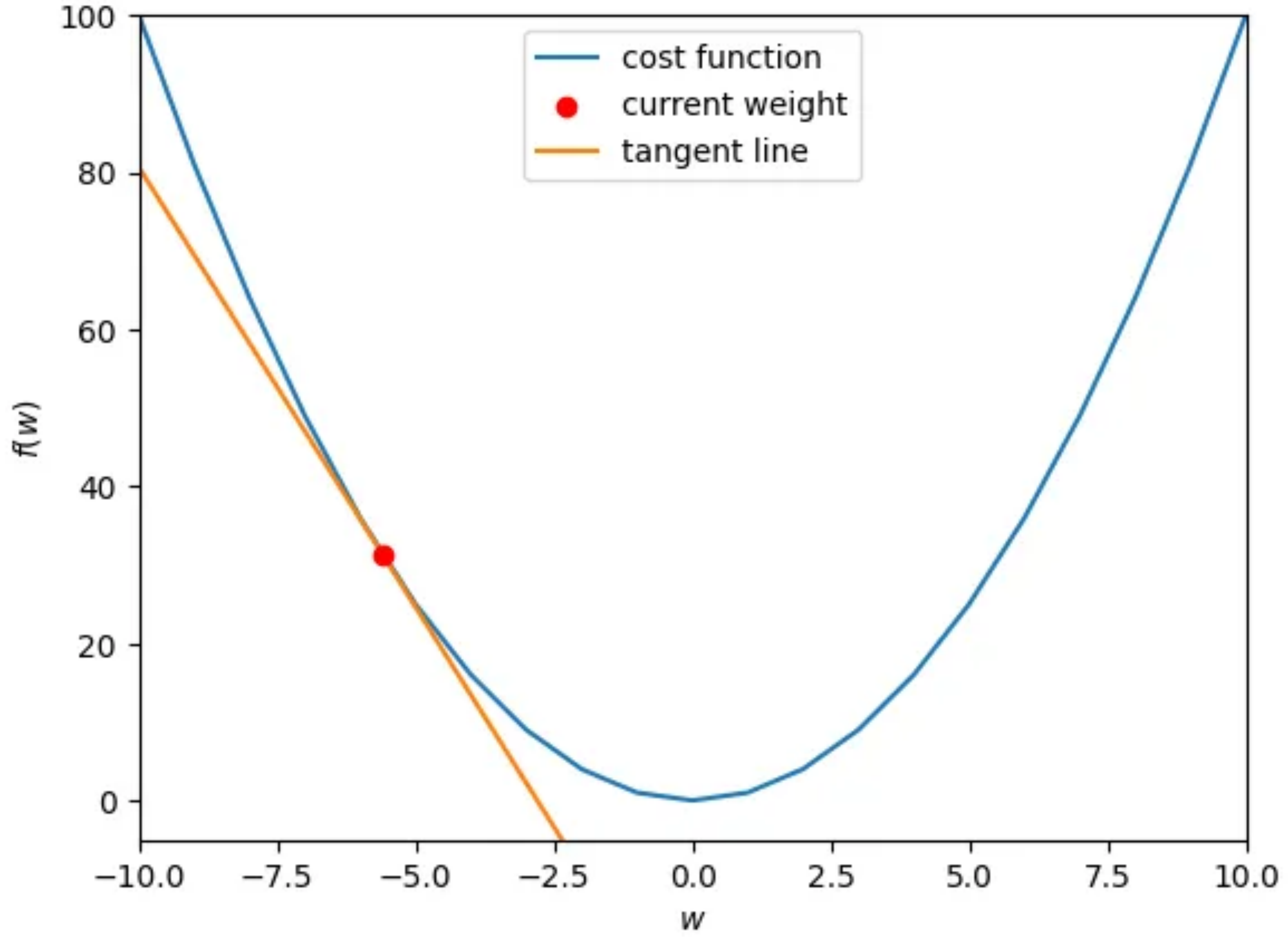
The cost function will be minimised (or loss is minimised) in iterations as parameters are **updated**

https://medium.com/@hunter-j-phillips/a-simple-introduction-to-gradient-descent-1f32a08b0deb#:~:text=By%20using%20partial%20derivatives%2C%20a,influence%20on%20the%20cost%20function.

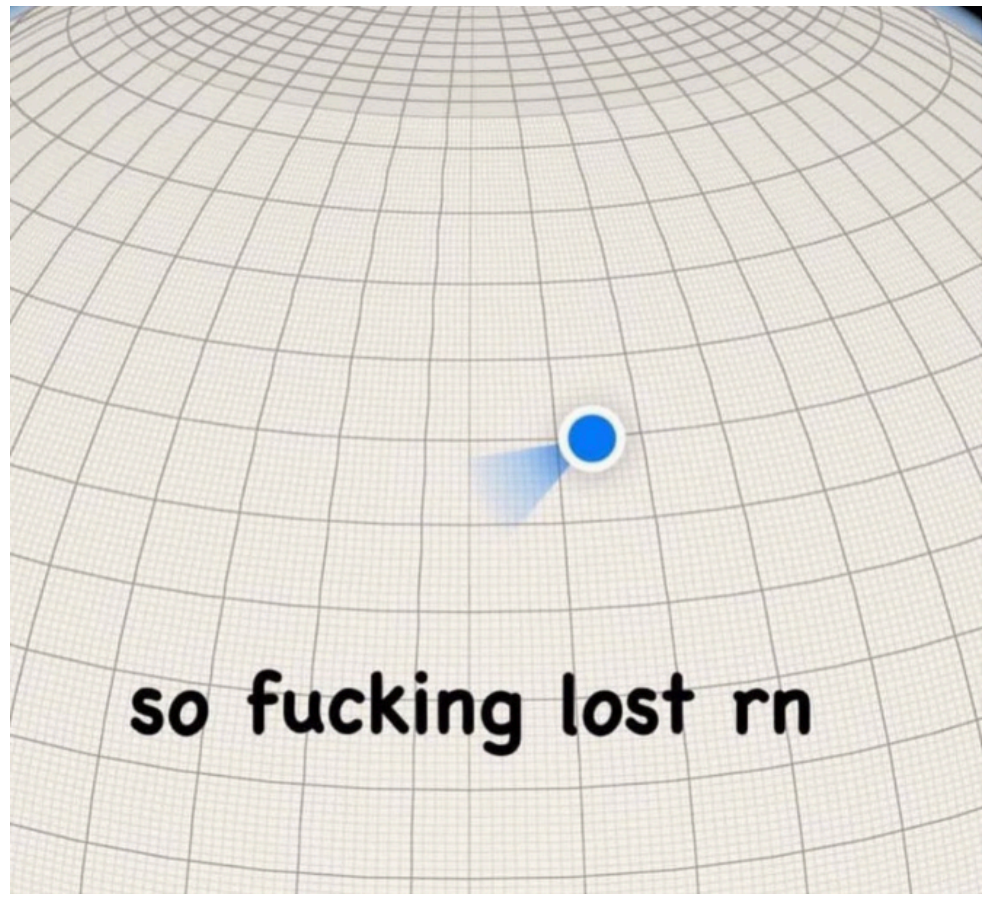
PARTIAL DERIVATIVES



A PARTIAL DERIVATIVE MEASURES HOW A FUNCTION CHANGES AS ONE OF ITS VARIABLES IS VARIED, HOLDING OTHERS CONSTANT.



PARTIAL DERIVATIVES



AN EXAMPLE

Consider this function f of two variables, x and y:

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

Let's find the partial derivative of f with respect to x.

We differentiate f with respect to x, **treating y as a constant.**

(Showing it on the board)

RULES FOR DIFFERENTIATION

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

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

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

$$\frac{d}{dx} b^x = b^x \ln(b)$$ Exponential Rule

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

Differentiation Rules

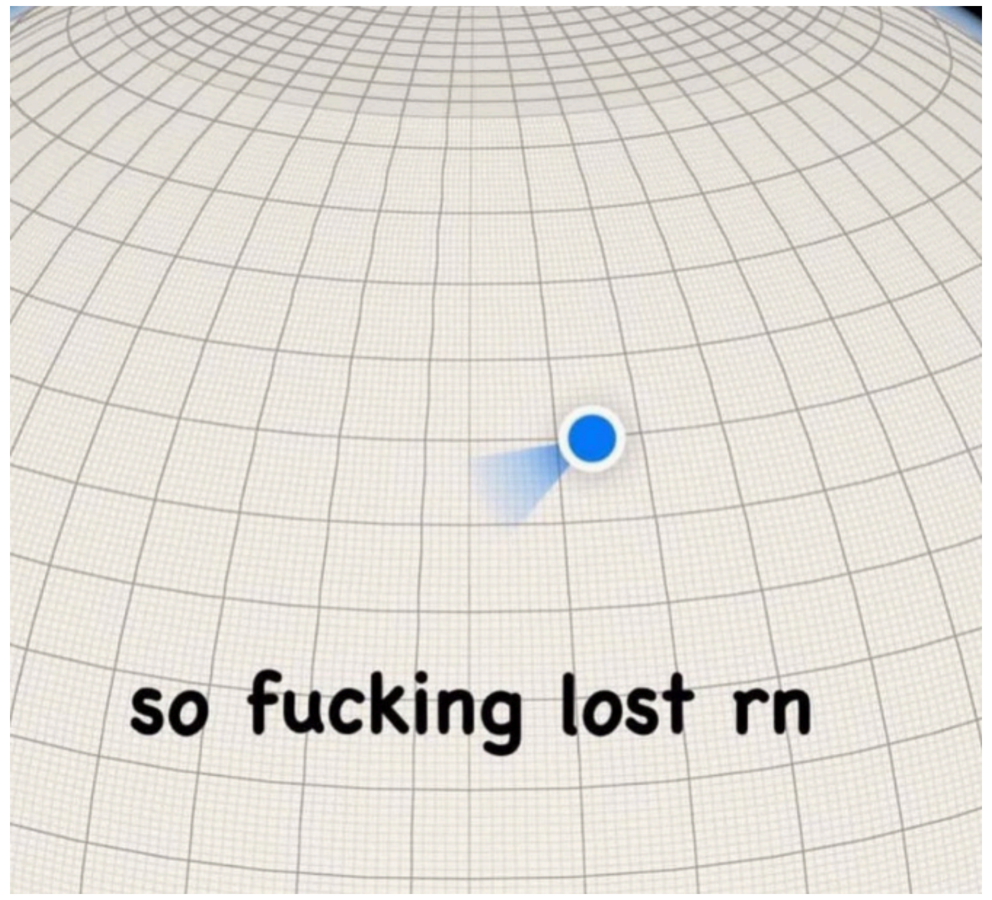
cuemath THE MATH EXPERT

Constant Rule	$\frac{d}{dx}[C] = 0$
Power Rule	$\frac{d}{dx} x^n = nx^{n-1}$
Product Rule	$\frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$
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	$\frac{d}{dx}[f(g(x))] = f'(g(x))\, g'(x)$

https://www.cuemath.com/calculus/differentiation/

https://study.com/academy/lesson/using-the-chain-rule-to-differentiate-complex-functions.html

PARTIAL DERIVATIVES



AN EXAMPLE

Consider this function f of two variables, x and y:

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

Let's find the partial derivative of f with respect to x.

We differentiate f with respect to x, **treating y as a constant.**

(Showing it on the board)

Should become: $$\frac{df}{dx} = 2xy + 3y$$

What does the partial derivative tell us then?

The partial derivative with respect to x, $\frac{df}{dx}$, tells us how f changes as x changes, keeping y constant

YOUR TURN: PARTIAL DERIVATIVES

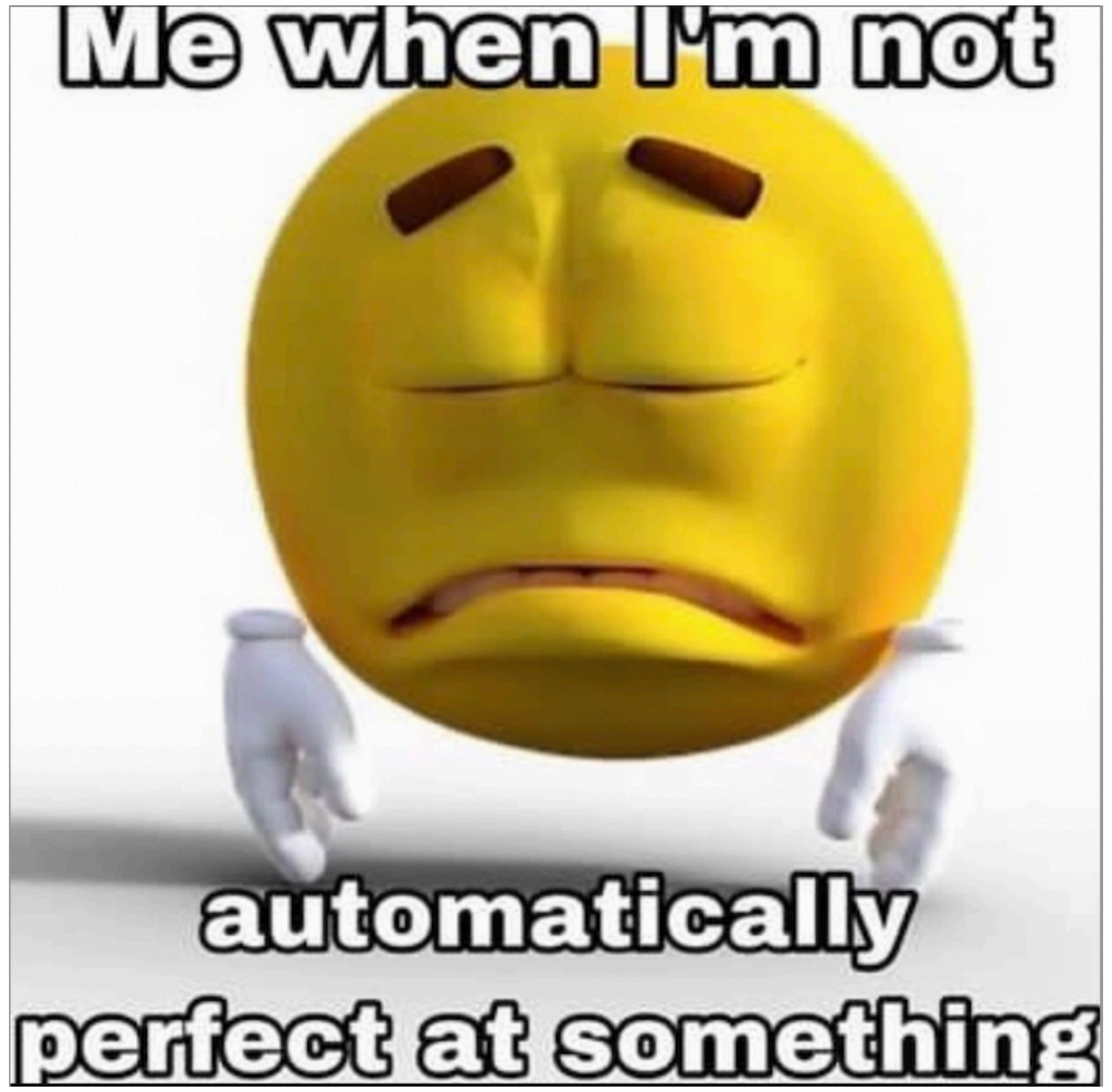
Consider this (same as before) function f of two variables, x and y:

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

In groups, find the partial derivative of f with respect to y.

We differentiate f with respect to y, **treating x as a constant this time.**

YOUR TURN: PARTIAL DERIVATIVES



Consider this (same as before) function f of two variables, x and y:

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

In groups, find the partial derivative of f with respect to y.

We differentiate f with respect to y, **treating x as a constant this time.**

(Showing it on the board)

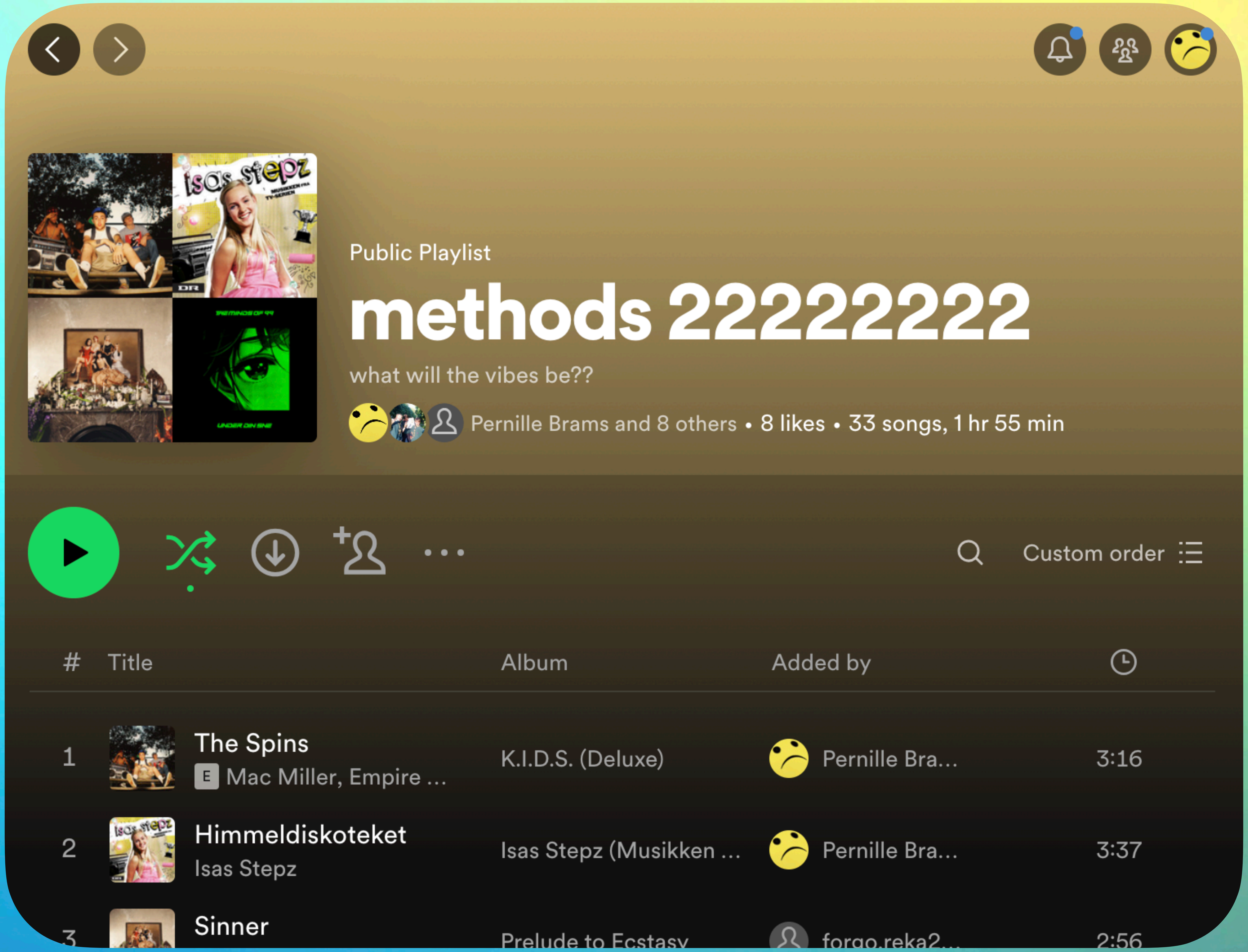
Should become:

$$\frac{df}{dy} = x^2 + 3x + 4y$$

What does the partial derivative tell us then?

The partial derivative with respect to y, $\frac{df}{dy}$, tells us how f changes as y changes, keeping x constant

BREAK

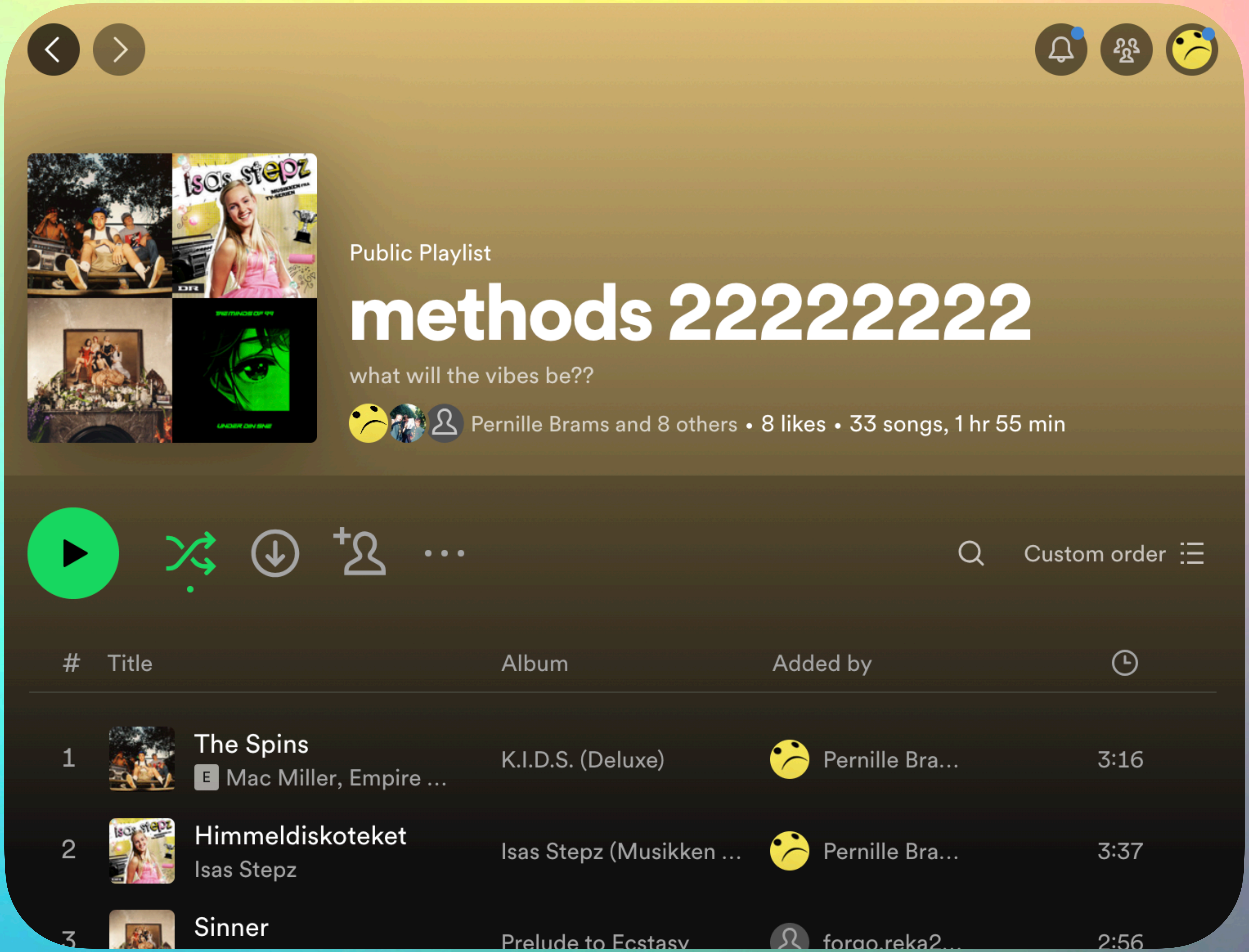


Collab: https://open.spotify.com/playlist/5UUiKD15vyFwymQ4qLur9V?si=cf80f6c72721427c&pt=a27065eeba53e23fe2dd160612aa5598

EXERCISES

In the GILL book:
6.4
6.7
6.9
6.1 (only do the first 4 and try to graph the functions in R)
As well as a couple of exercises in the notebook.

If you have more time. Finish exercises from review exercises or do more exercises from chapter 6 (for instance 6.11).



Collab: https://open.spotify.com/playlist/5UUiKD15vyFwymQ4qLur9V?si=cf80f6c72721427c&pt=a27065eeba53e23fe2dd160612aa5598