GENERAL ASSEMBLY

FEATURE SCALING

Sam Stack, Ritika Bhasker

AGENDA

- What are the benefits of feature scaling?
- Identify situations where feature scaling is beneficial.
- Scale data using Python and SKLearn.

GENERAL ASSEMBLY

WHAT IS SCALING?

SCALING

SCALING

SCALING

- A set of techniques used to counteract the effects of different features having different scales or ranges.

SCALING

Why would we want to scale data?

SCALING

- We scale our data to prevent a particular feature from having a disproportionate impact on our model.
- Gradient descent also converges much faster with feature scaling.

SCALING

- We DON'T use feature scaling with algorithms that consider features independently (tree-based models).
- We DO use feature scaling for models that rely heavily on distances to make predictions (the majority of classifiers).

HOW DO WE SCALE OUR DATA

- **STANDARDISATION**

  Centers (centres ☹) the data around 0 and scales it with respect to the standard deviation of the feature.

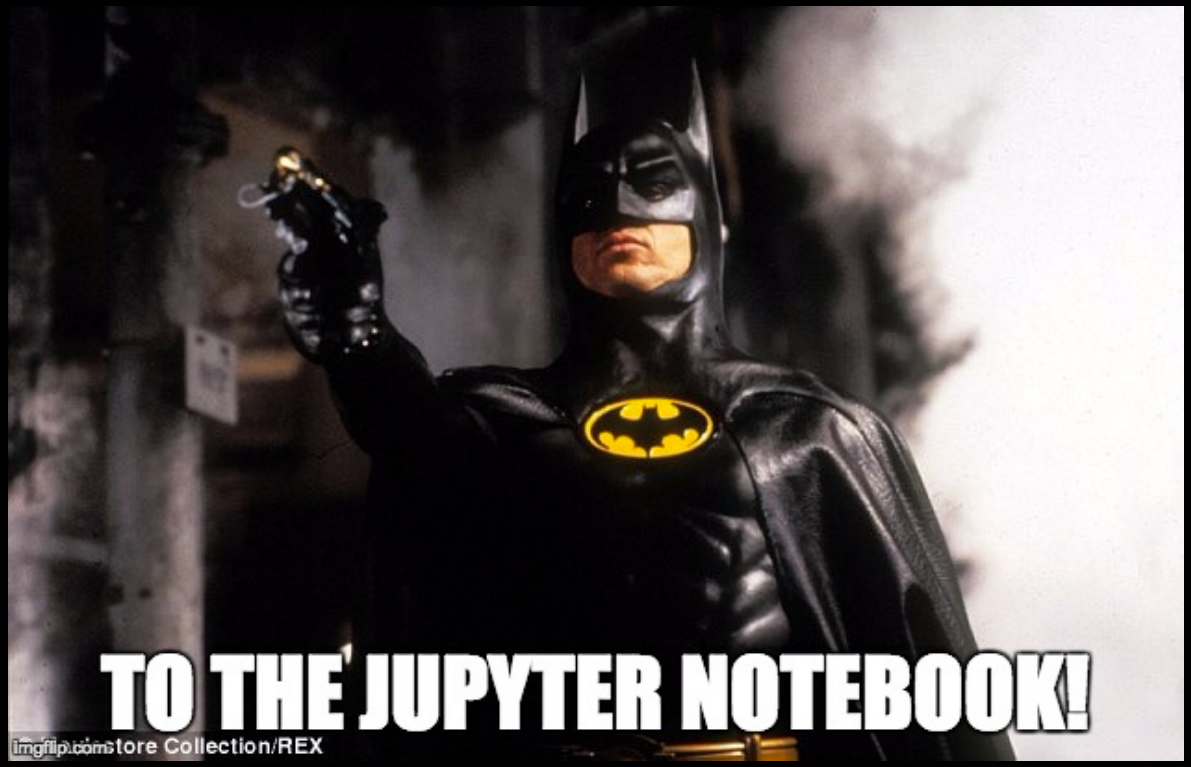
$$x_{new} = \frac{x - \mu}{\sigma}$$

HOW DO WE SCALE OUR DATA

- **NORMALISATION**

  Scales the variable so that the minimum is zero and the maximum is 1

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}$$


TO THE JUPYTER NOTEBOOK!
imgflip.com
store Collection/REX