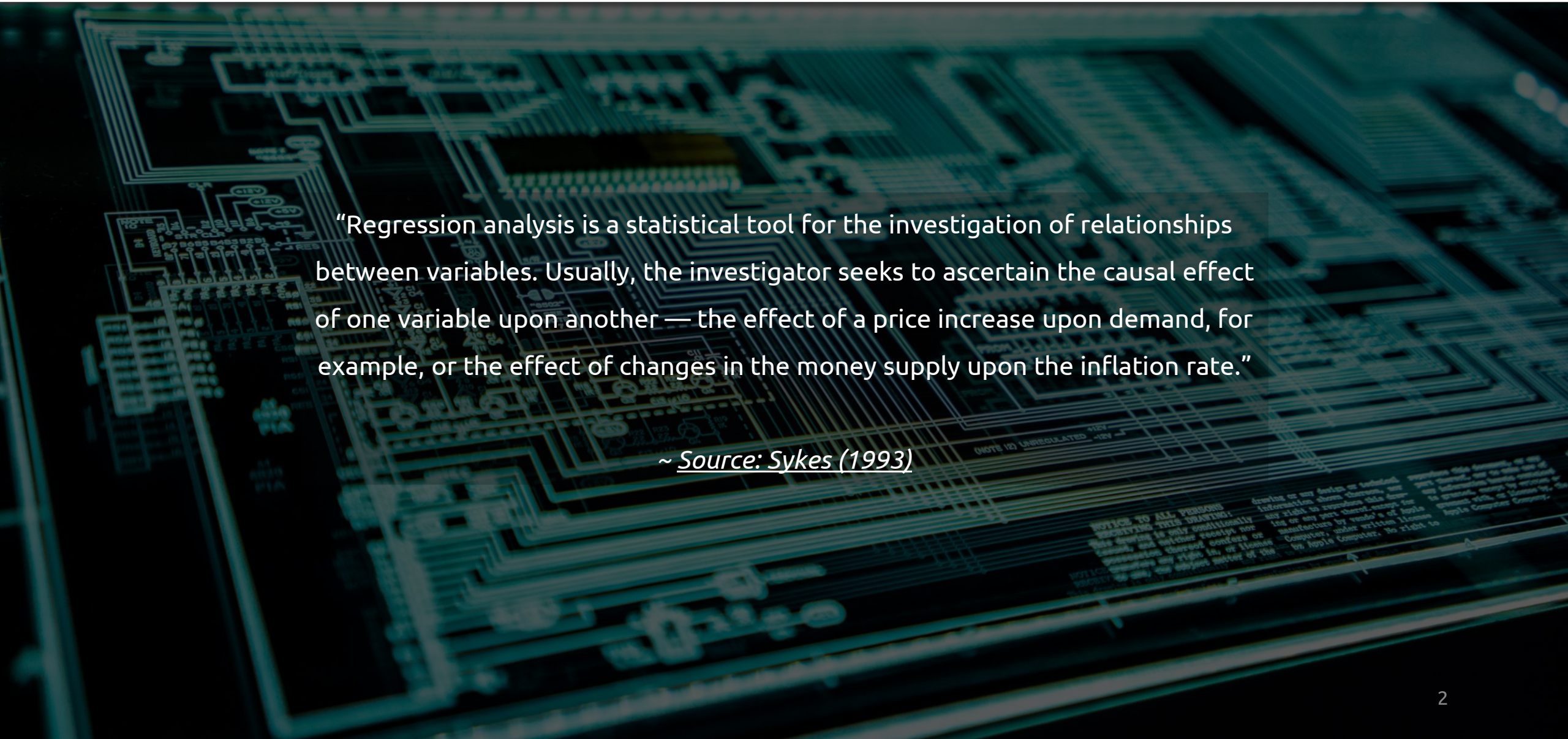
Regression



"Regression analysis is a statistical tool for the investigation of relationships between variables. Usually, the investigator seeks to ascertain the causal effect of one variable upon another — the effect of a price increase upon demand, for example, or the effect of changes in the money supply upon the inflation rate."

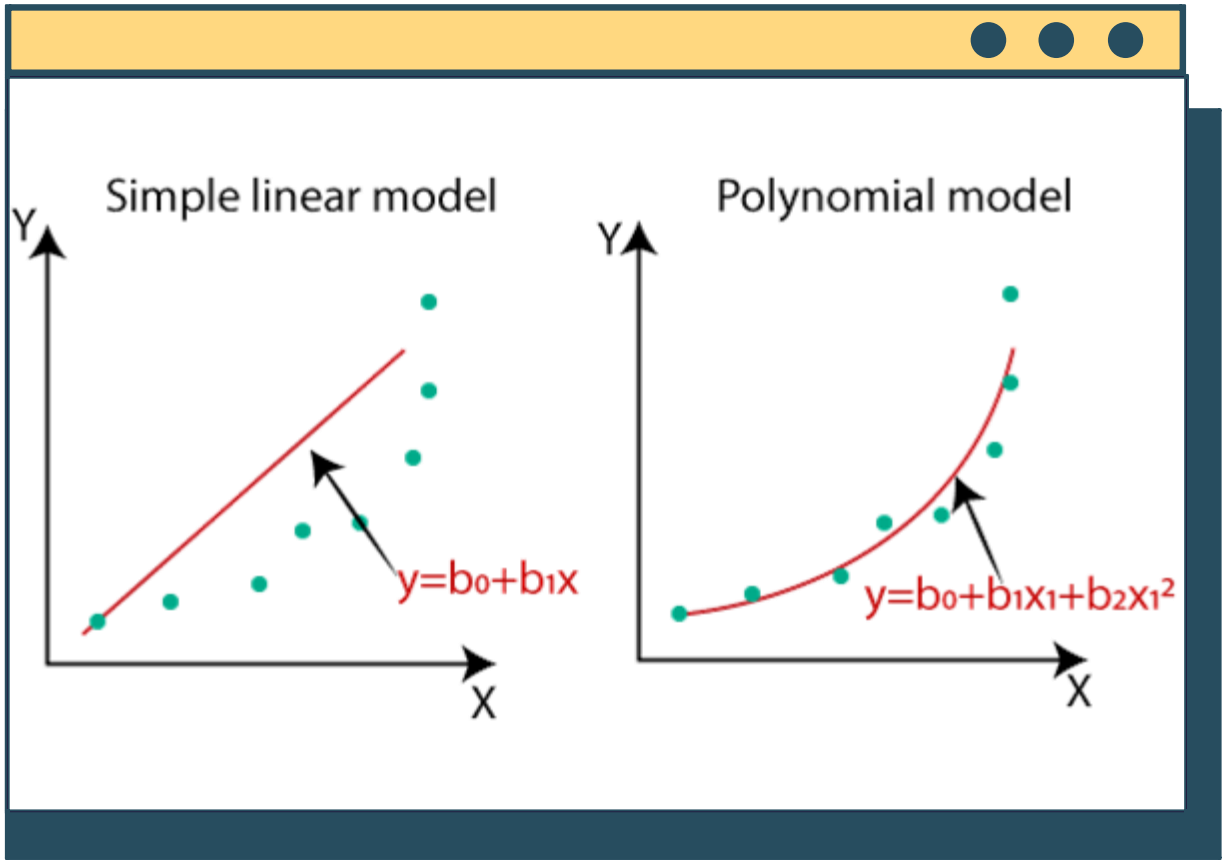
~ *Source: Sykes (1993)*

REGRESSION

Regression in machine learning consists of mathematical methods that allows **predictions** of a **continuous outcome (y)** based on the value of one or more **predictor variables (x)**.

Keyword: Continuous Data

TYPES OF REGRESSION

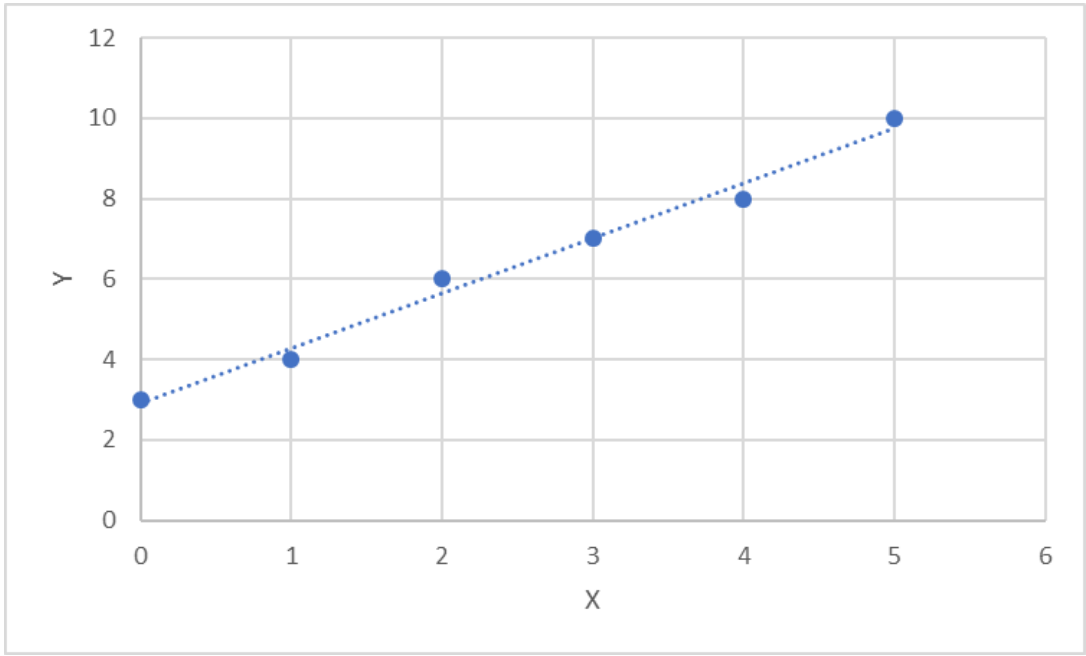


LINEAR REGRESSION

Linear regression finds the linear relationship between the dependent variable and one or more independent variables using a best-fit straight line.

How to obtain the best fit line?
The best-fit line is obtained using least square method by reducing the sum of the squares of the offsets (residual part/error) of the points from the curve, i.e. minimizing the error.

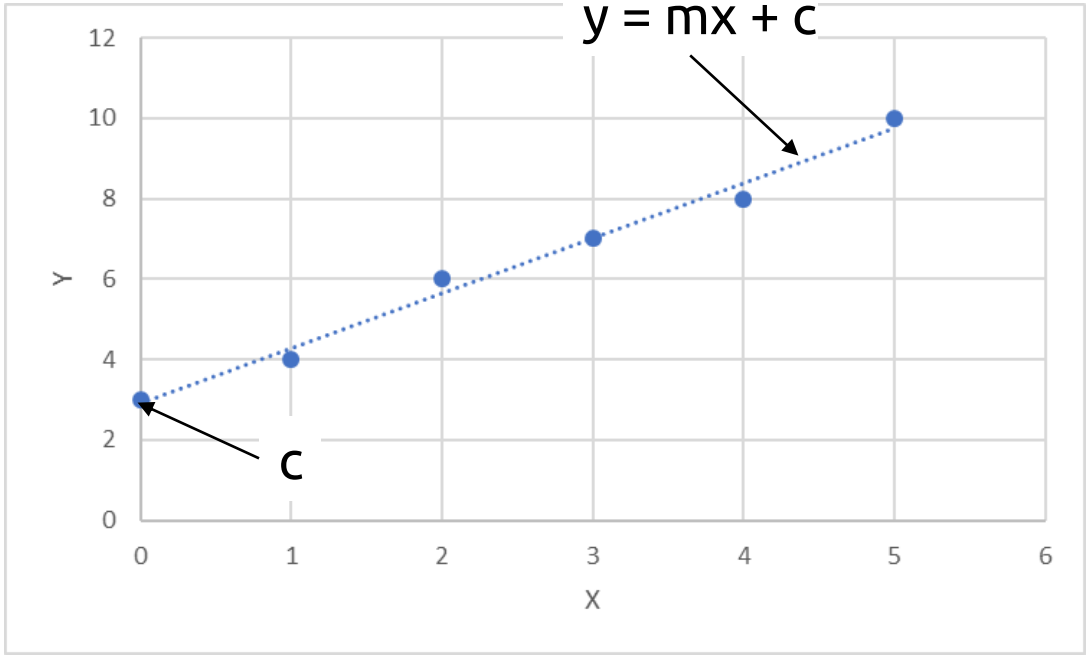
Error/loss is calculated by subtracting the actual value from the predicted one. Since the result from subtracting might be negative, we square the difference to make it a positive value.



LINEAR REGRESSION

$$y = mx + c$$

$y = dependent\ variable$
$m = the\ slope\ of\ the\ line$
$x = independent\ variable$
$c = y - intercept$



$$m = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(x_i - \bar{x})^2} \qquad c = \bar{y} - m\bar{x}$$

To help us get a better understanding let's simulate the process of finding best fit line intuitively from:
https://phet.colorado.edu/sims/html/least-squares-regression/latest/least-squares-regression_en.html

Thanks!

This is the end of Regression, see you in the next topic.