Implementing Simple Regression Models in Python

Vitthal Srinivasan
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Build regression models in Python

Make an informed choice of technology between Python, R and Excel

Understand how Python views regression as a machine learning problem

Ease of Prototyping

Python is not as good as Excel or R for prototyping

Robustness and Re-use

But it is significantly better for use in production-style environments

“Make the common use-case easy and the difficult use-case possible.”

Regression: Excel, R or Python?

Excel

Create a regression slide for an important presentation

R

Create a regression case study for a seminar

Python

Build a trading model that scrapes websites, combines sentiment analysis and regression

Regression: Excel, R or Python?

Excel

Presentations

R

Seminars

Python

Trading models

Regression is a Great Tool

Powerful

Perfectly suited to two common use-cases

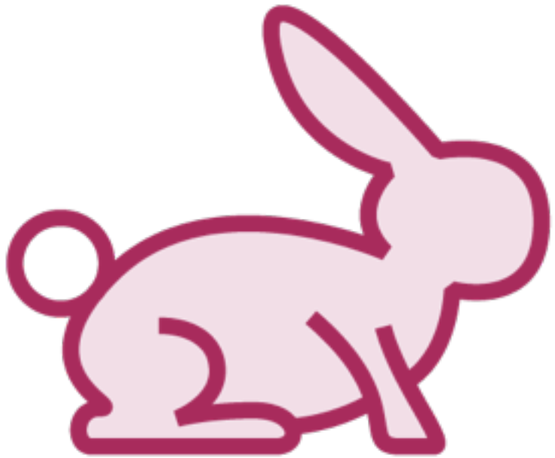
Versatile

Easily extended to non-linear relationships

Deep

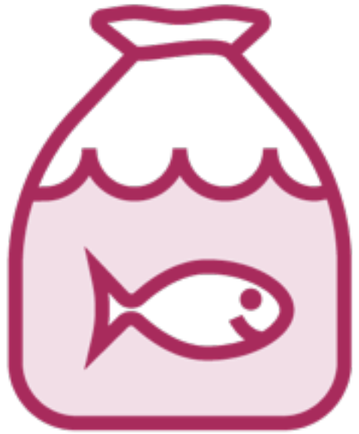
The first “crossover hit” from Machine Learning

Whales: Fish or Mammals?



Mammals

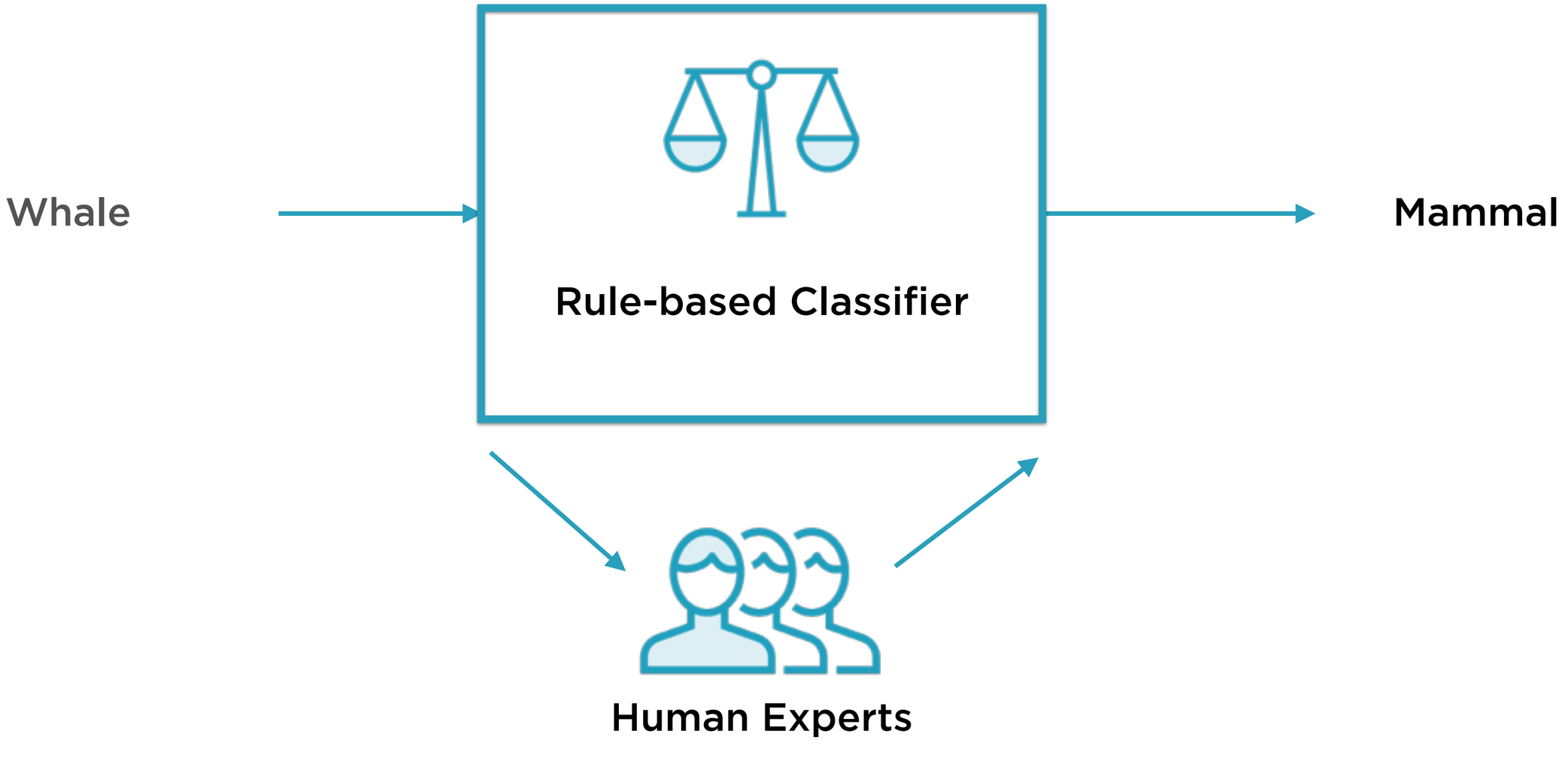
Members of the infraorder *Cetacea*



Fish

Look like fish, swim like fish, move with fish

Rule-based Binary Classifier



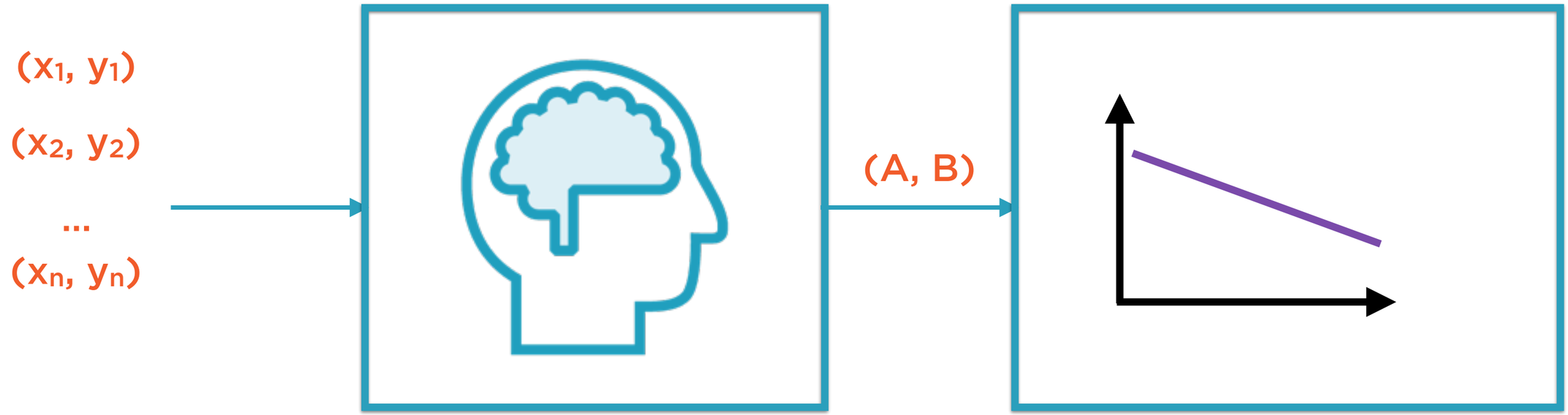
ML-based Binary Classifier

Corpus

Classification Algorithm

ML-based Classifier

ML-based Regression Model

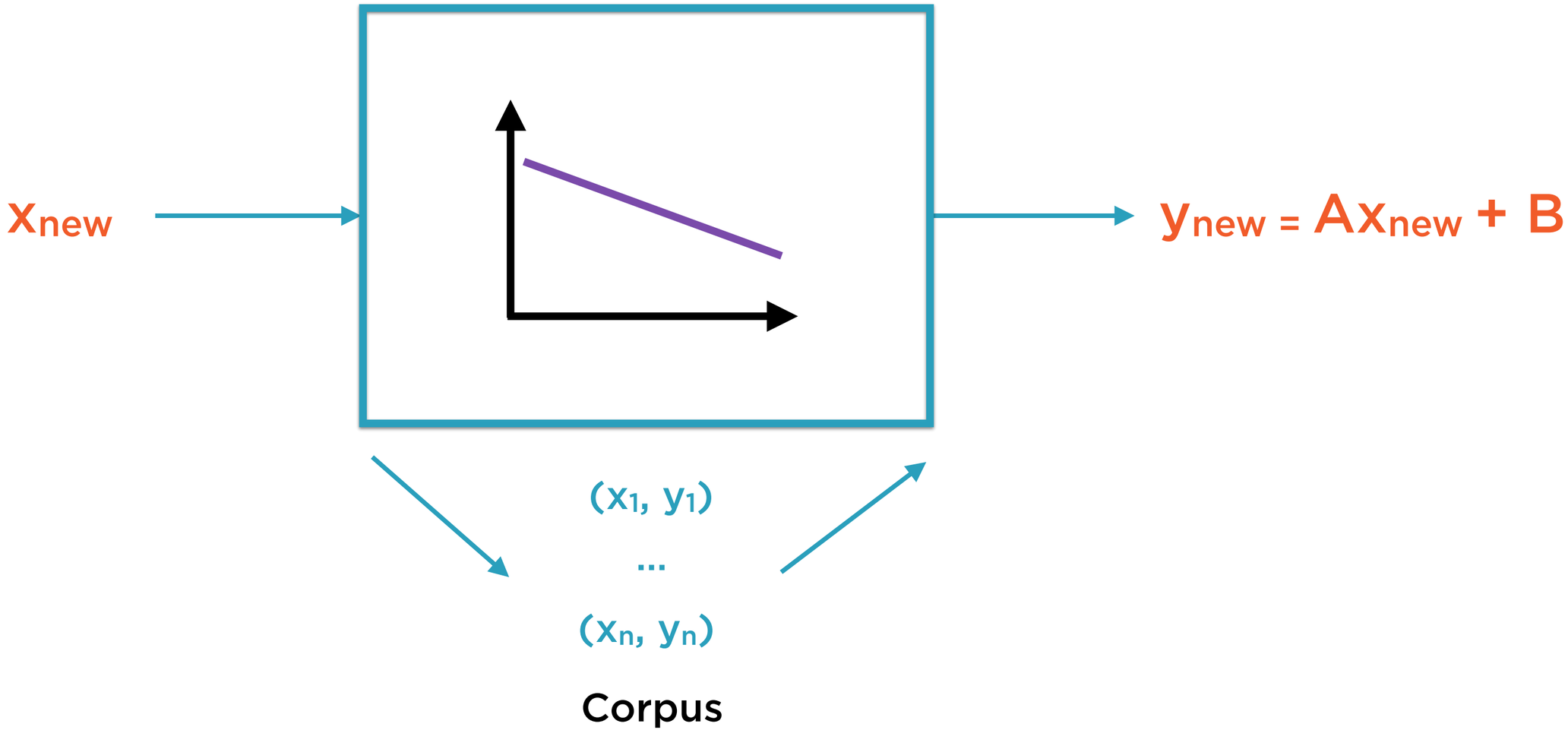


Corpus

Regression Algorithm

Regression Line: y = A + Bx

ML-based Regression Model



Demo

Simple regression models in Python

A Dream Team of Libraries

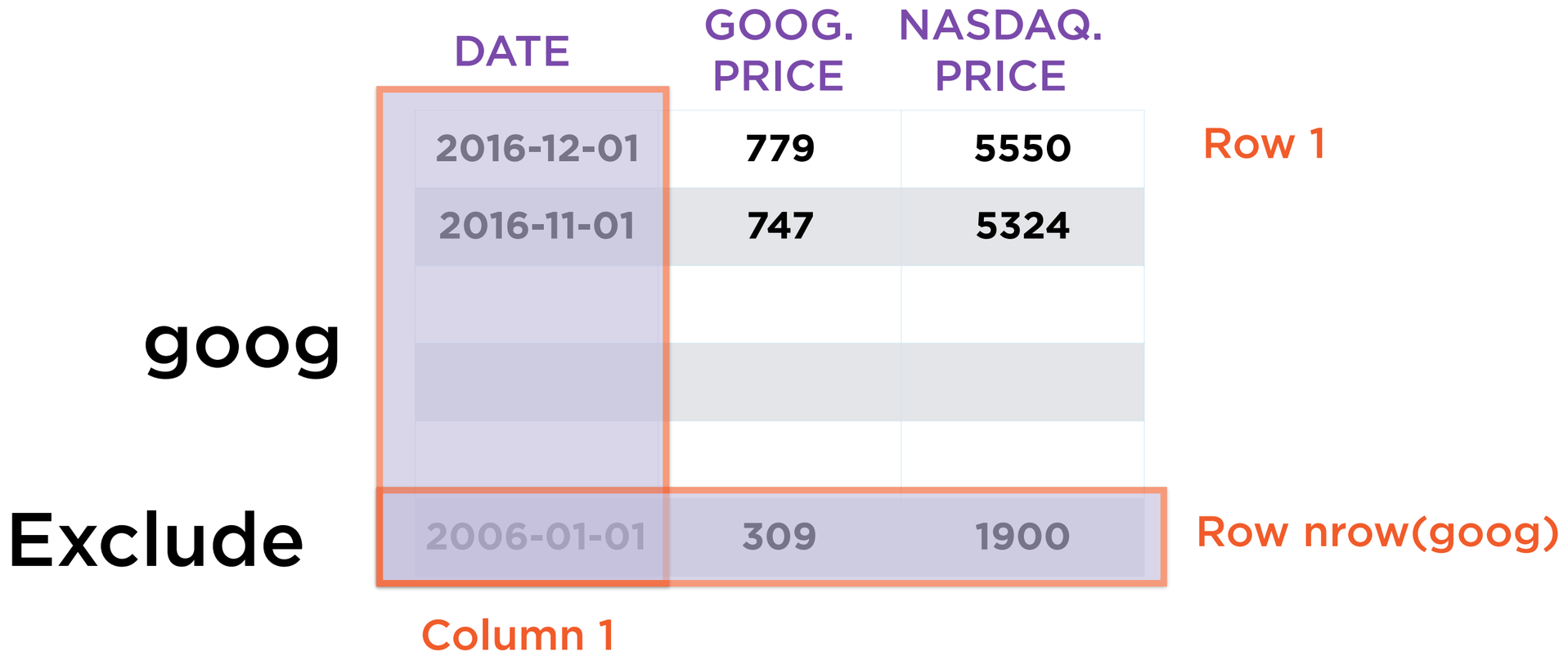
Pandas for dataframes

NumPy for arrays

Scikit for regression

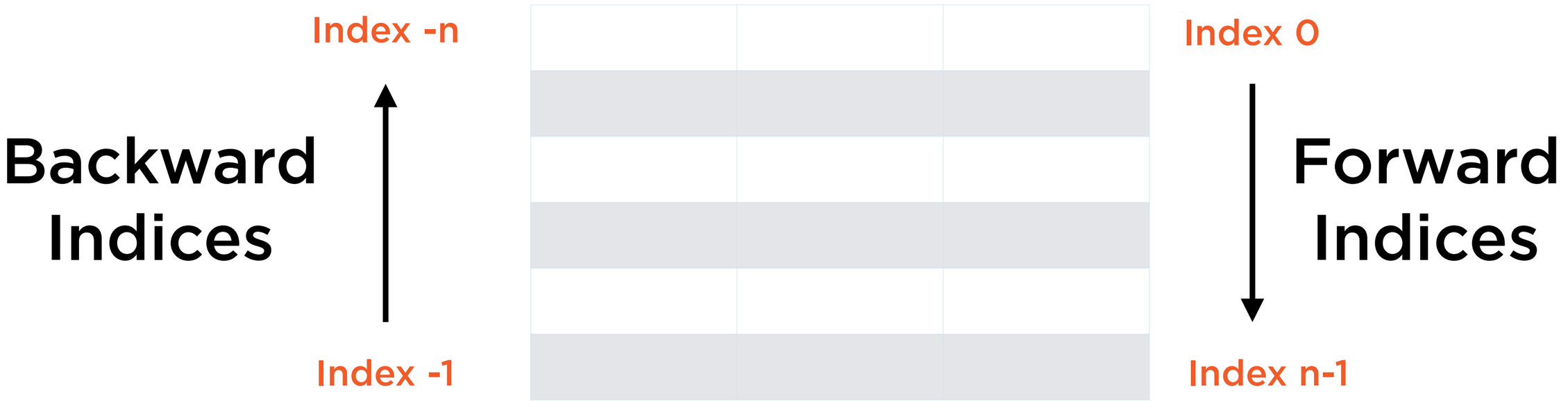
Matplotlib for plots

Negative Indices in R

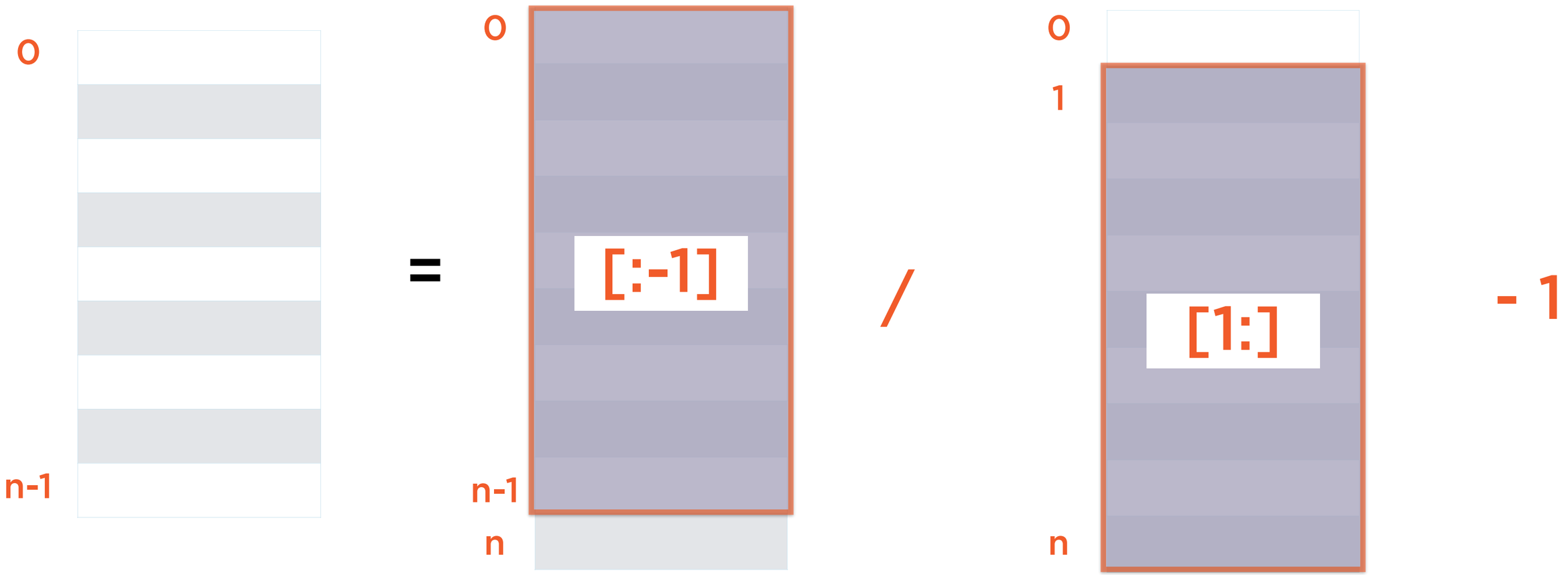



```
goog[-nrow(goog),-1]
```

Negative Indices In Python

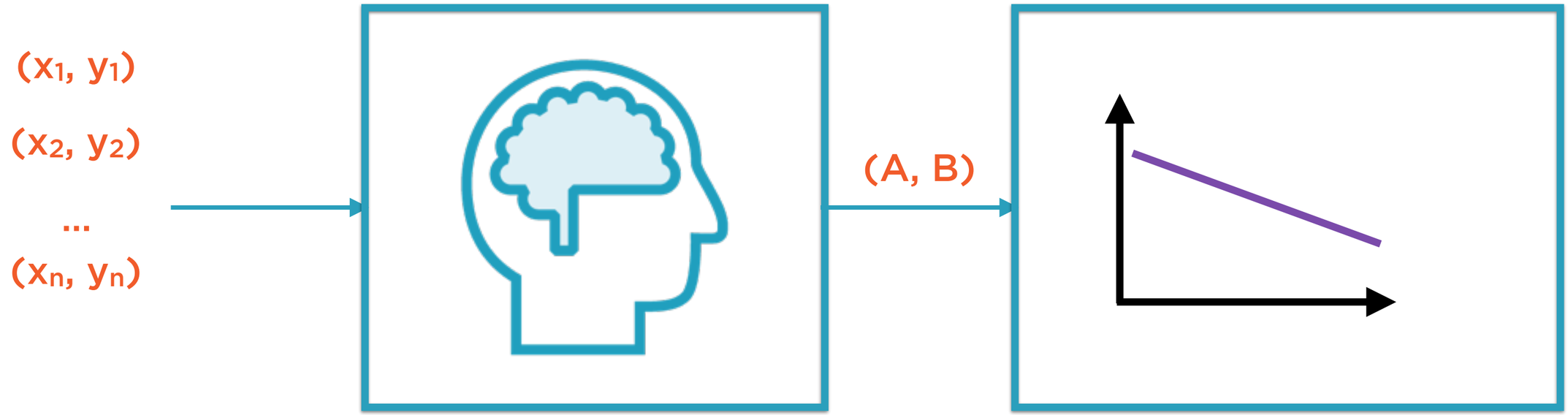


Prices to Returns



Returns = Prices[:-1] / Prices[1:] - 1

ML-based Regression Model



Corpus

Regression Algorithm

Regression Line: y = A + Bx

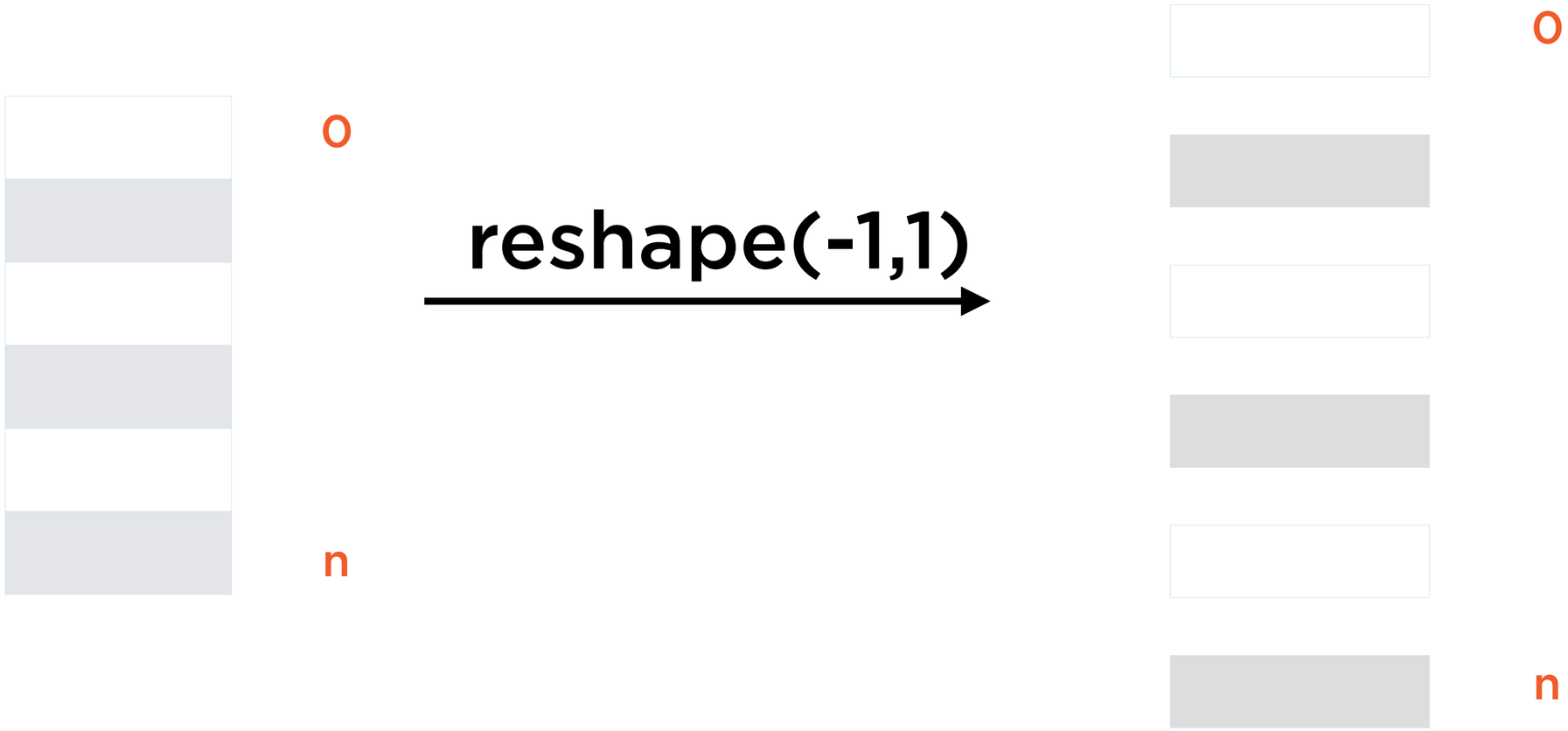
ML-based Regression Model

Corpus

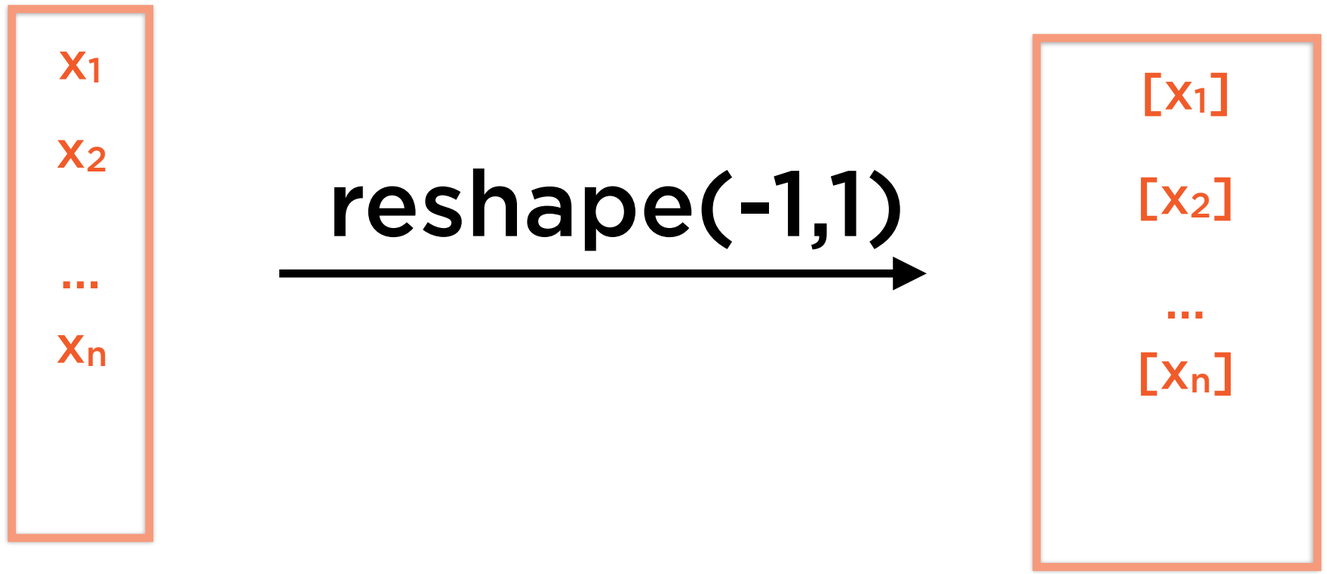
NumPy Linear Regression

Regression Line:
y = A + Bx

Reshaping in NumPy



Reshaping in NumPy



Summary

Understood when Python is the most appropriate tool for regression

Built regression models in Python