Implementing Logistic Regression Models in Excel

Vitthal Srinivasan
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Set up a logistic regression to predict whether a stock will rise or fall

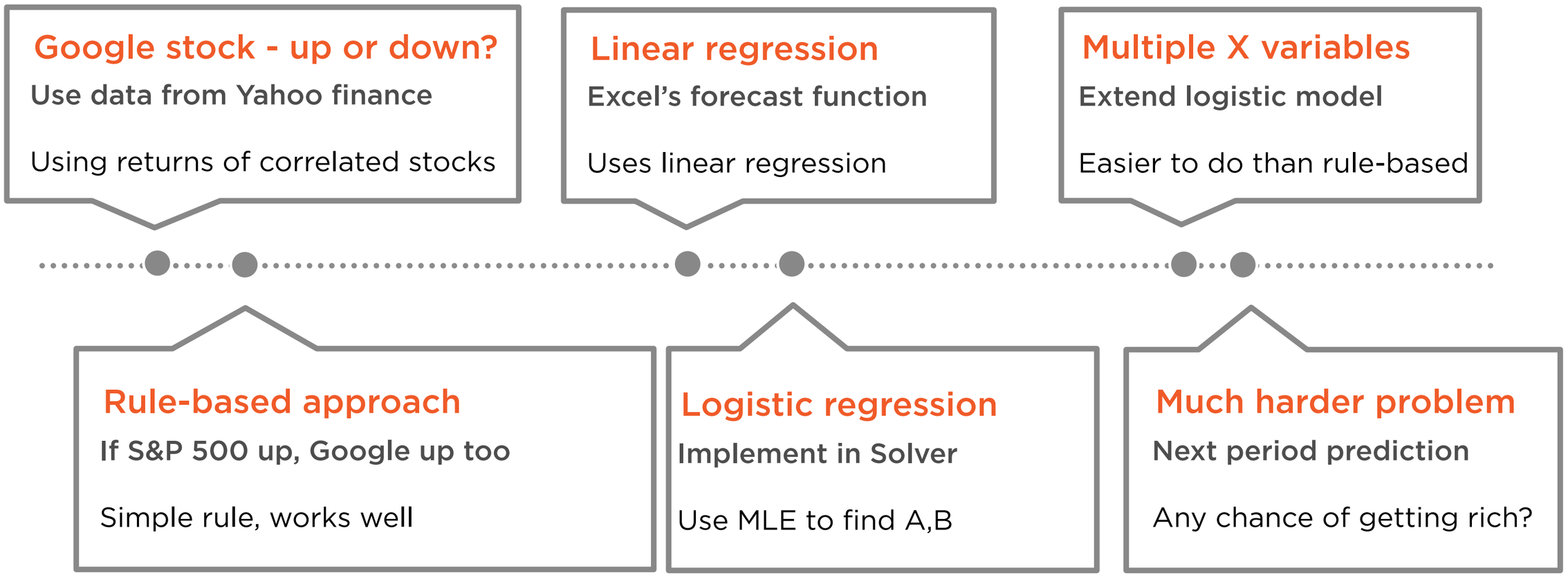
Solve this logistic regression in Excel using Solver

Contrast this solution to a rule-based approach

Extend the logistic regression to include multiple explanatory variables

Attempt a far more difficult task - predicting future returns using logistic regression

Logistic Regression in Excel



Demo

Implement Logistic Regression in Excel

Logistic Regression in Excel

Google stock - up or down?

Use data from Yahoo finance

Using returns of correlated stocks

Set up the Problem

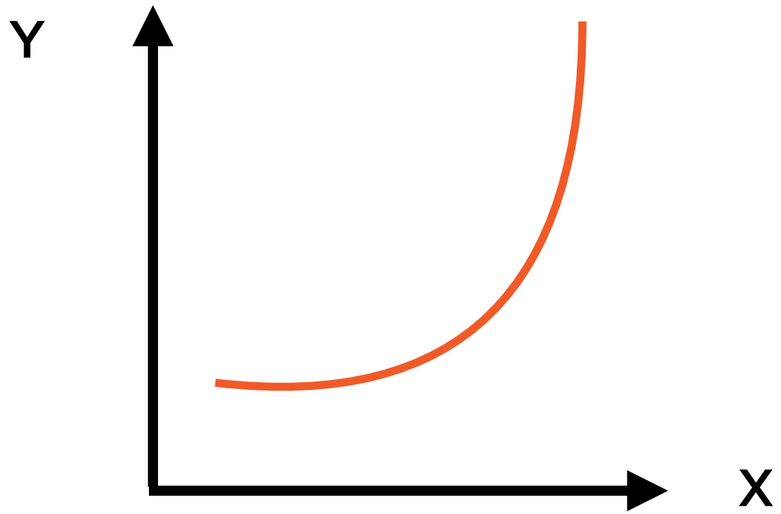
Cause

Changes in S&P 500

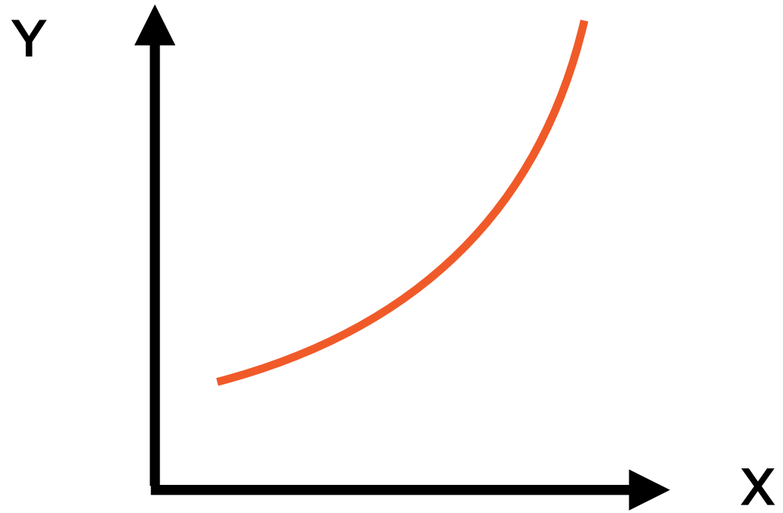
Effect

Changes in price of Google Stock

Never Regress Non-Stationary Data



Exponential



Polynomial

Smoothly trending data will lead to poor quality regression models

Convert Prices to Returns

$y'_{12} = \log y_2 - \log y_1$

$x'_{12} = \log x_2 - \log x_1$

Regress y' and x'

Log Differences

$y'_{12} = (y_2 - y_{1)}/y_1$

$x'_{12} = (x_2 - x_{1)}/x_1$

Regress y' and x'

Returns

Take first differences of smooth data converting either to log differences or returns

Set up the Problem

y = Returns on Google stock (GOOG)

x = Returns on S&P 500 (S&P500)

Set up the Problem

DATE	GOOG	S&P500
2017-02-01	813.67	2316.10
2017-01-01	796.79	2278.87
2005-01-01	97.71	1181.27

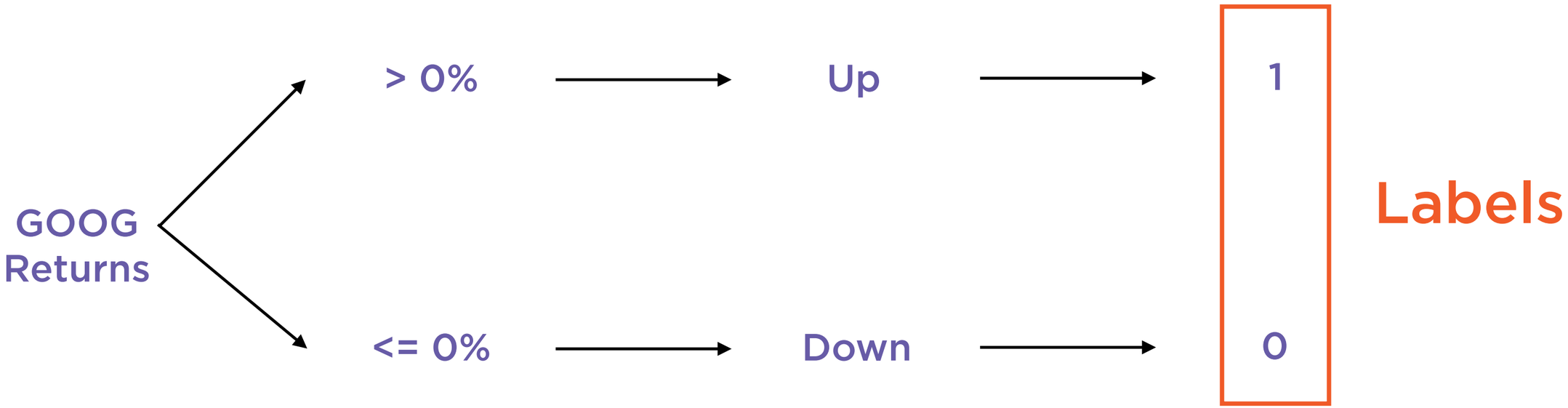
Download prices from Yahoo finance

Set up the Problem

DATE	GOOG	S&P500
2005-01-01	97.71	1181.27
2017-01-01	796.79	2278.87
2017-02-01	813.67	2316.10

Sort date from oldest to newest to calculate returns

Set up the Problem



Label GOOG returns as binary (1,0)

Logistic Regression in Excel

Google stock - up or down?

Use data from Yahoo finance

Using returns of correlated stocks

Rule-based approach

If S&P 500 up, Google up too

Simple rule, works well

Rule-based Approach

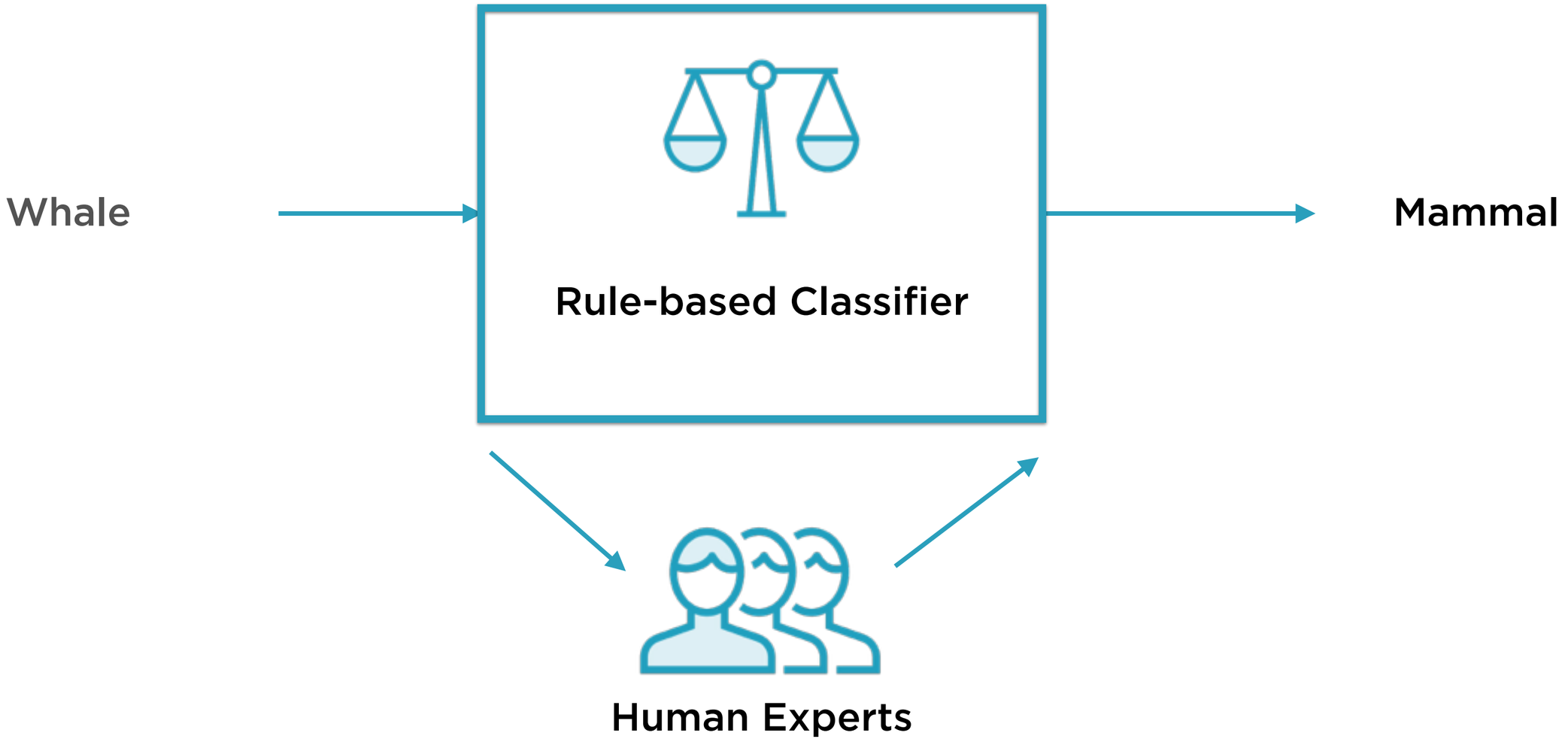
Cause

Changes in S&P 500

Effect

Changes in price of Google Stock

Rule-based Binary Classifier



Rule-based Approach

x = Returns on S&P 500 (S&P500)

y = Returns on Google stock (GOOG)

Rule: If S&P500 is up, then GOOG will be up too

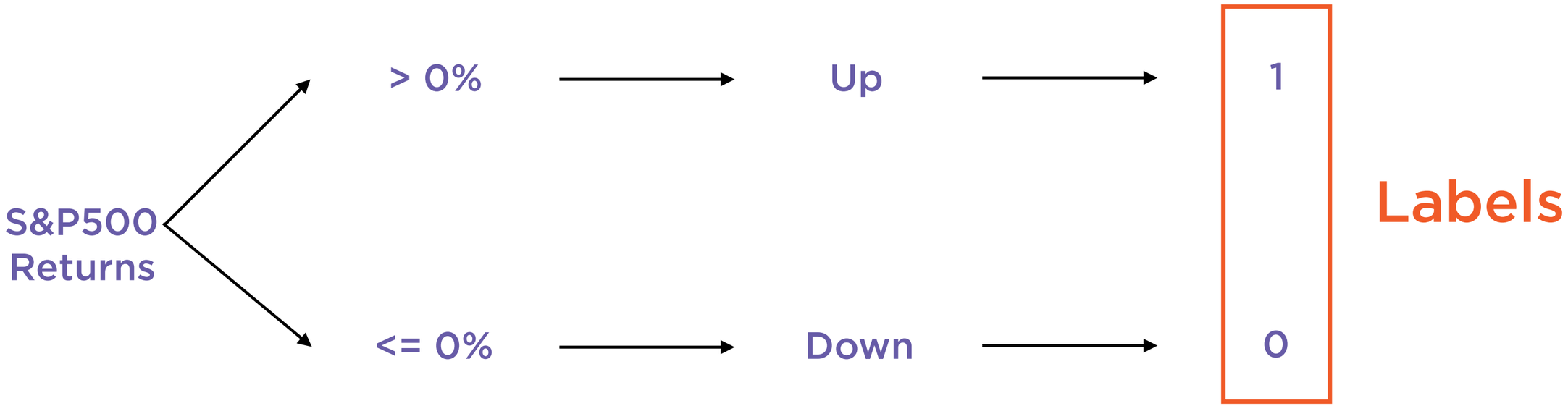
Rule-based Approach

x = Returns on S&P 500 (S&P500)

y = Returns on Google stock (GOOG)

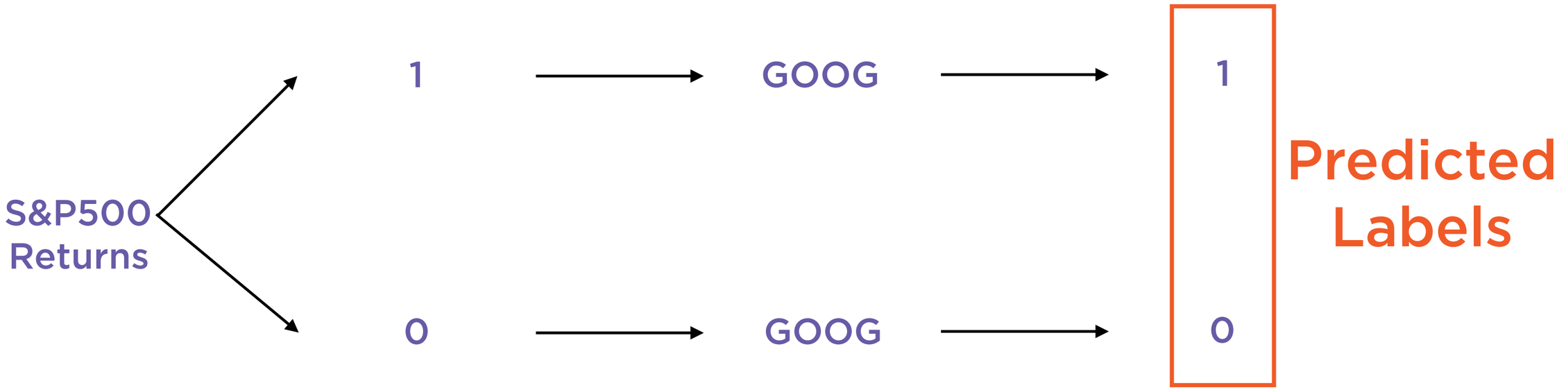
Rule: If S&P500 is down, then GOOG will be down too

Rule-based Approach



Label S&P500 returns as binary (1,0)

Rule-based Approach



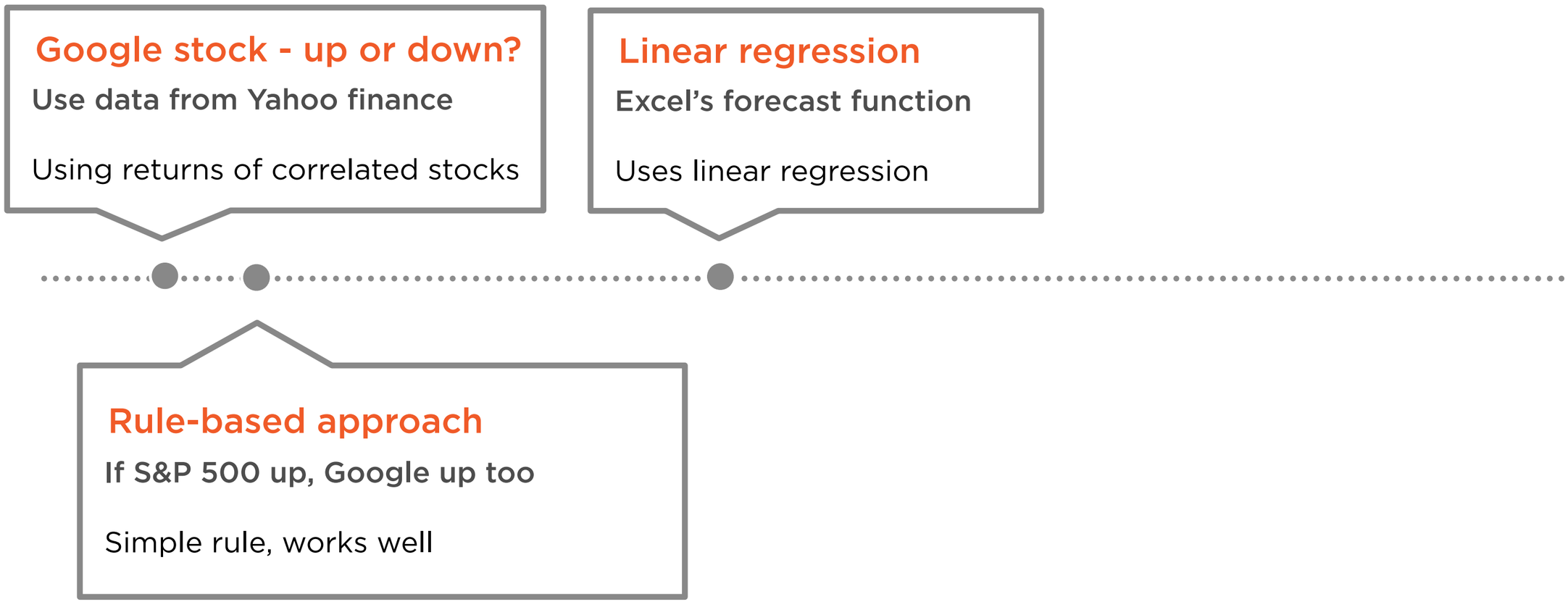
Apply our rule and assign S&P500's labels to GOOG

Rule-based Approach

DATE	ACTUAL	PREDICTED
2005-01-01	NA	NA
2005-02-01	0	1
2005-03-01	0	0
2017-01-01	1	1
2017-02-01	1	1

Compare GOOG's actual labels vs. predicted labels

Logistic Regression in Excel



Odds from Probabilities

$$\text{Odds}(p) = \frac{p}{1-p}$$

Odds of an Event

$$p = \frac{1}{1 + e^{-(A+Bx)}}$$

$$p = \frac{e^{A + Bx}}{1 + e^{A + Bx}}$$

$$1 - p = \frac{1}{1 + e^{A + Bx}}$$

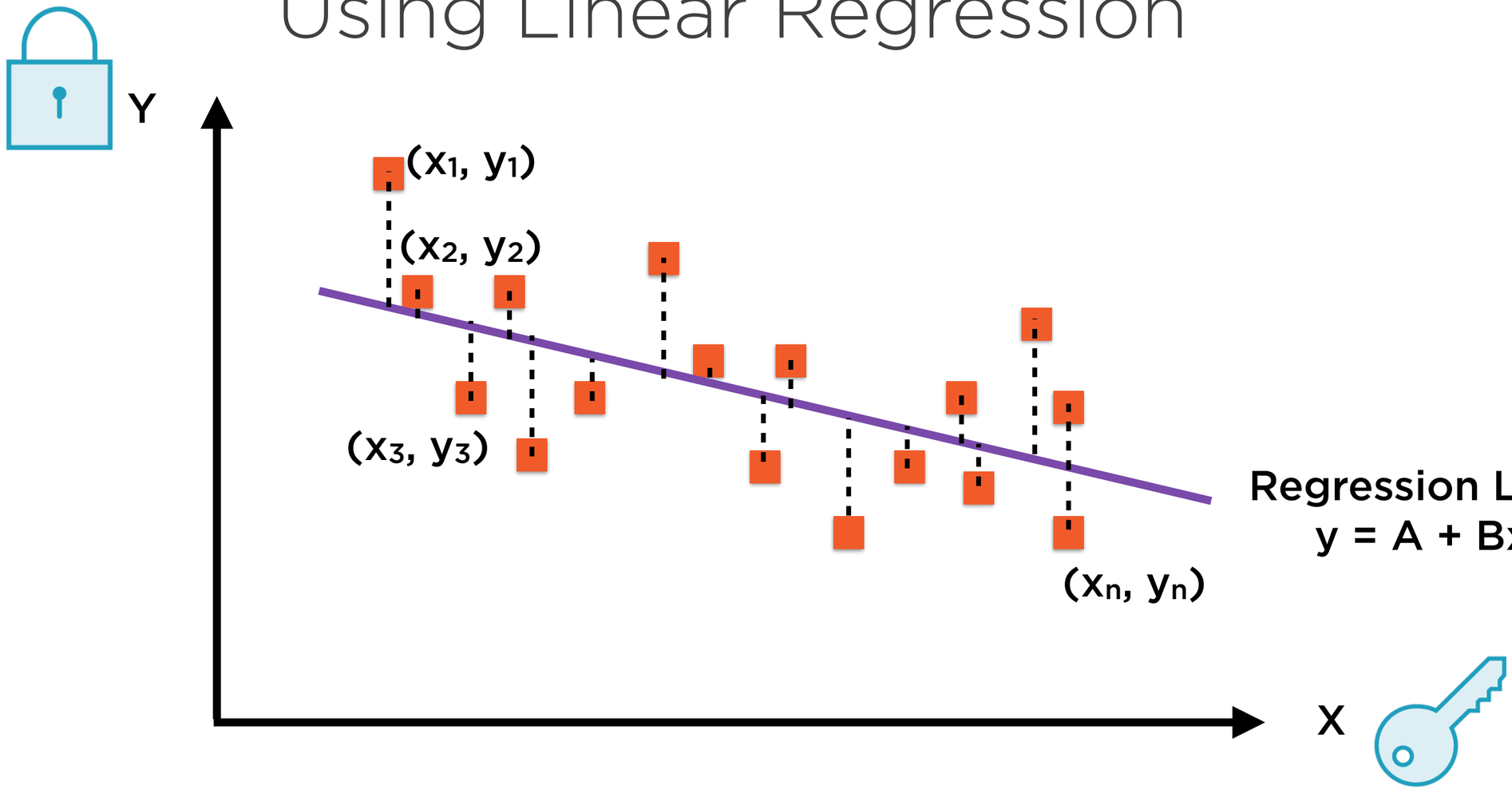
Odds of an Event

$$p = \frac{e^{A + Bx}}{1 + e^{A + Bx}}$$

$$1 - p = \frac{1}{1 + e^{A + Bx}}$$

$$\text{Odds}(p) = \frac{p}{1 - p} = e^{A + Bx}$$

Using Linear Regression



Represent all n points as
(x_i, y_i), where i = 1 to n

Using Linear Regression

```
y=FORECAST(x,known_y's,known_x's)
```

y = Returns on Google stock for current month

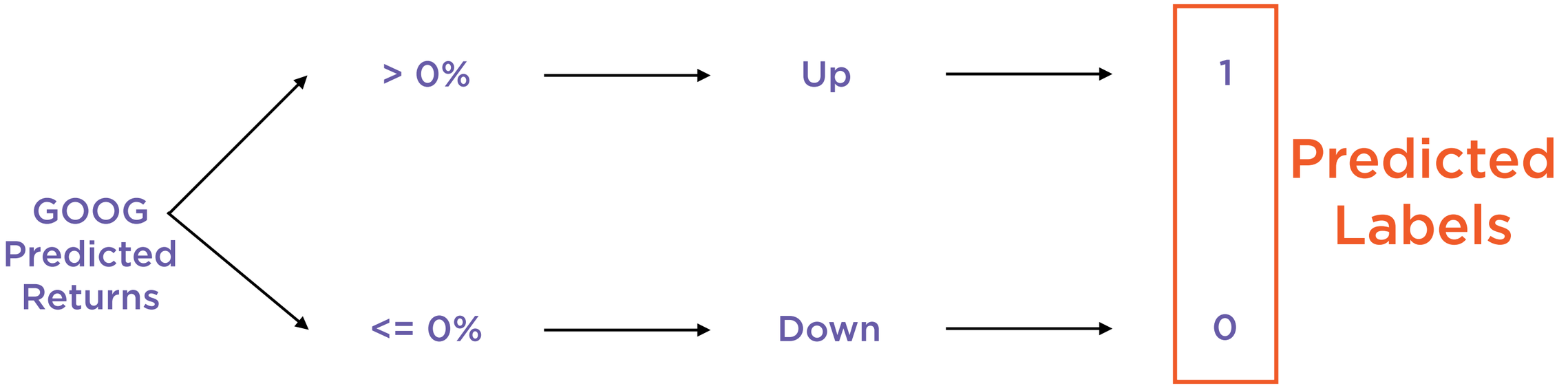
x = Returns on S&P 500 for current month

known_y's = Returns on Google stock for all months

known_x's = Returns on S&P 500 for all months

Predict GOOG's returns using forecasting

Using Linear Regression



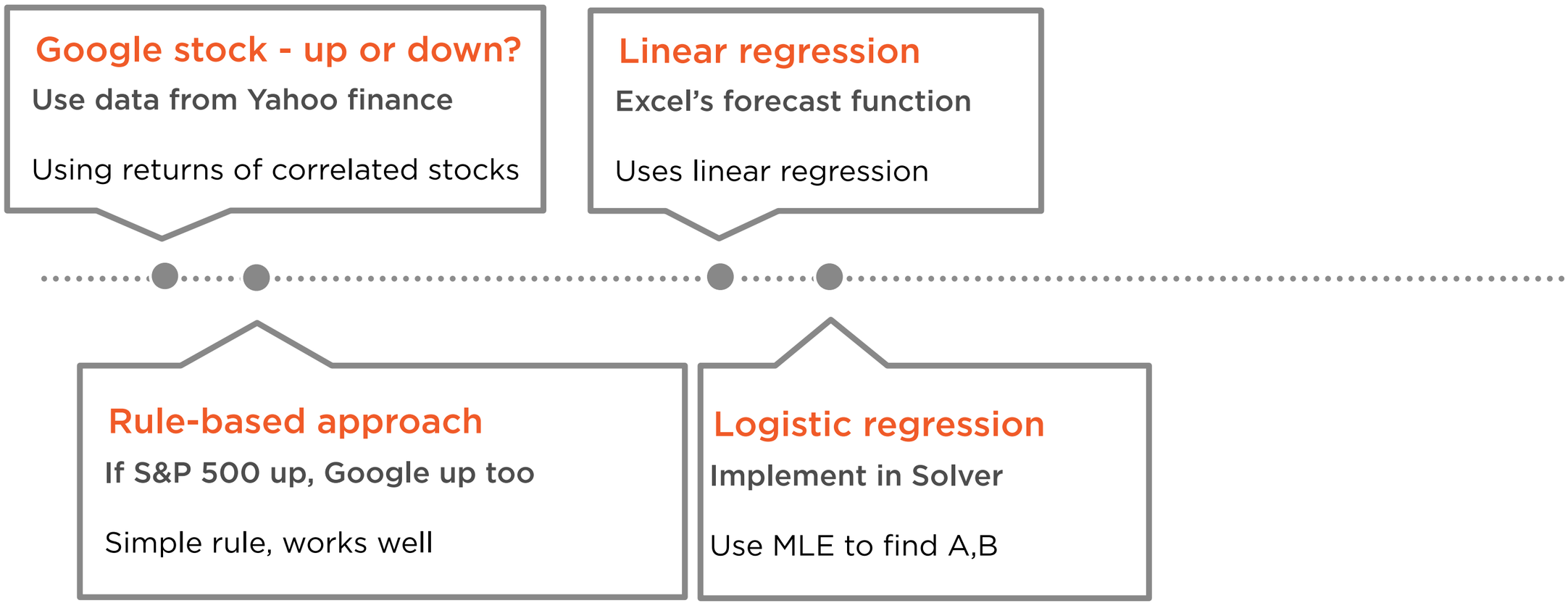
Label GOOG predicted returns as binary (1,0)

Using Linear Regression

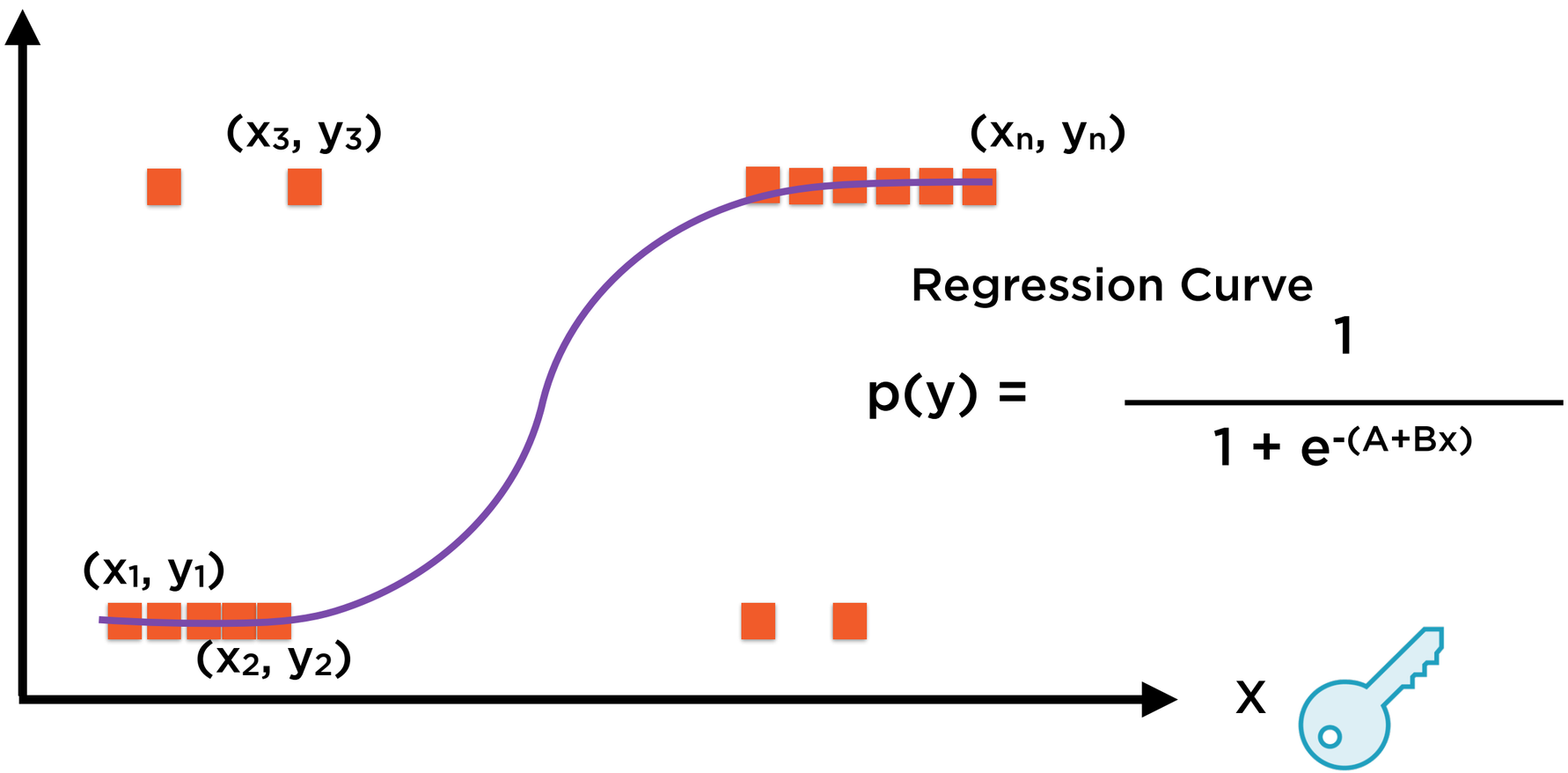
DATE	ACTUAL	PREDICTED
2005-01-01	NA	NA
2005-02-01	0	1
2005-03-01	0	0
2017-01-01	1	1
2017-02-01	1	1

Compare GOOG's actual labels vs. predicted labels

Logistic Regression in Excel



Using Logistic Regression



Represent all n points as
(x_i, y_i), where i = 1 to n

Using Logistic Regression

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

P(y) = Probability of Google going up in the current month i

x = Returns on S&P 500 for current month

Use logistic regression to find probabilities (assuming A, B = 0)

Using Logistic Regression

$$LL = \ln L = \sum_{i=1}^{n} [y_i \ln(p_i) + (1-y_i)\ln(1-p_i)]$$

Calculate the log likelihood

Logistic Regression

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

Solve for A and B that “best fit” the data

Using Logistic Regression

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

Use Excel's Solver to calculate A, B, maximizing the log likelihood

Using Logistic Regression

Excel options → **Add-ins** → **Solver Add-in**

Install Solver as an add-in to Excel

Using Logistic Regression

Set target cell

Choose a function (max., min., value)

Choose cells to change

Define constraints

Solver parameters

Using Logistic Regression

Log Likelihood	Maximize
A, B in logistic regression equation	None

Keying in solver parameters

Using Logistic Regression

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

Solver gives the values of A, B while maximizing the log likelihood

Rule-based or ML-based?

ML-based	Rule-based
Dynamic	Static
Experts optional	Experts required
Corpus required	Corpus optional
Training step	No training step

Using Logistic Regression

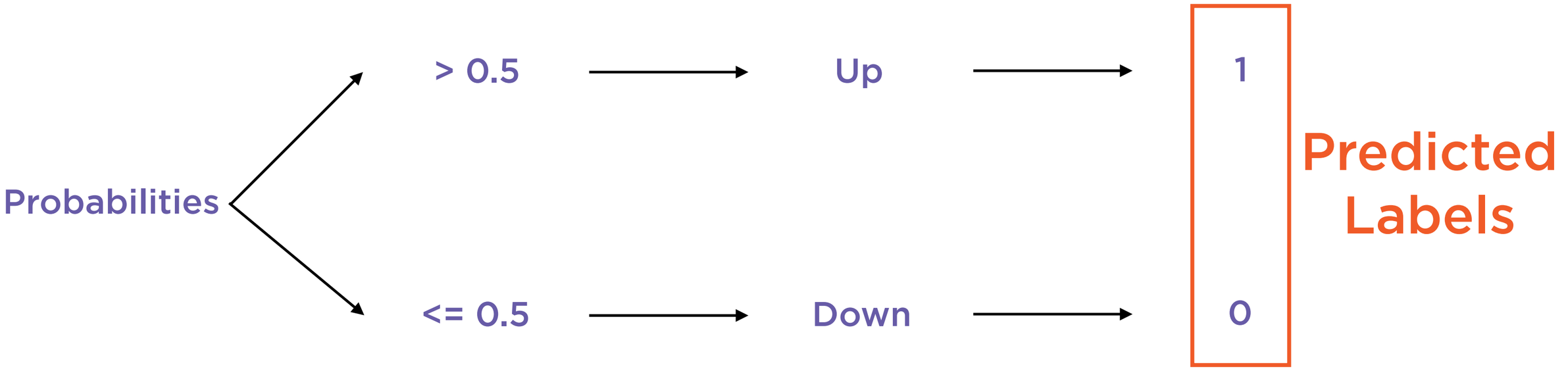
$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

P(y) = Probability of Google going up in the current month i

x = Returns on S&P 500 for current month

Recalculate probabilities, using the new values of A, B

Using Logistic Regression



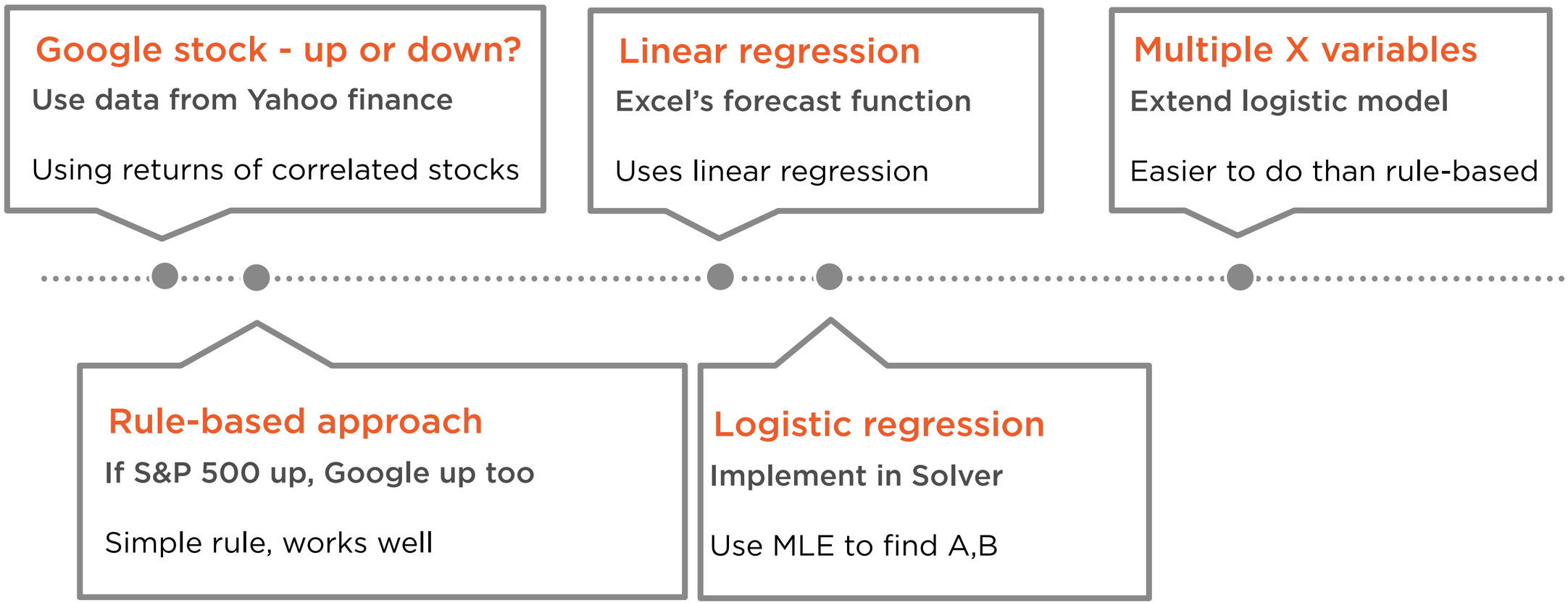
Label probabilities from the logistic regression as binary (1,0)

Using Logistic Regression

DATE	ACTUAL	PREDICTED
2005-01-01	NA	NA
2005-02-01	0	1
2005-03-01	0	0
2017-01-01	1	1
2017-02-01	1	1

Compare GOOG's actual labels vs. predicted labels of probabilities from logistic regression

Logistic Regression in Excel



Using Logistic Regression

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

P(y) = Probability of Google going up in the current month i

x = Returns on S&P 500 for current month

Recalculate probabilities, using the new values of A, B

Using Logistic Regression

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

P(y) = Probability of Google going up in the current month i

x = Returns on S&P 500 for current month

Recalculate probabilities, using the new values of A, B

Multiple X Variables - Easy

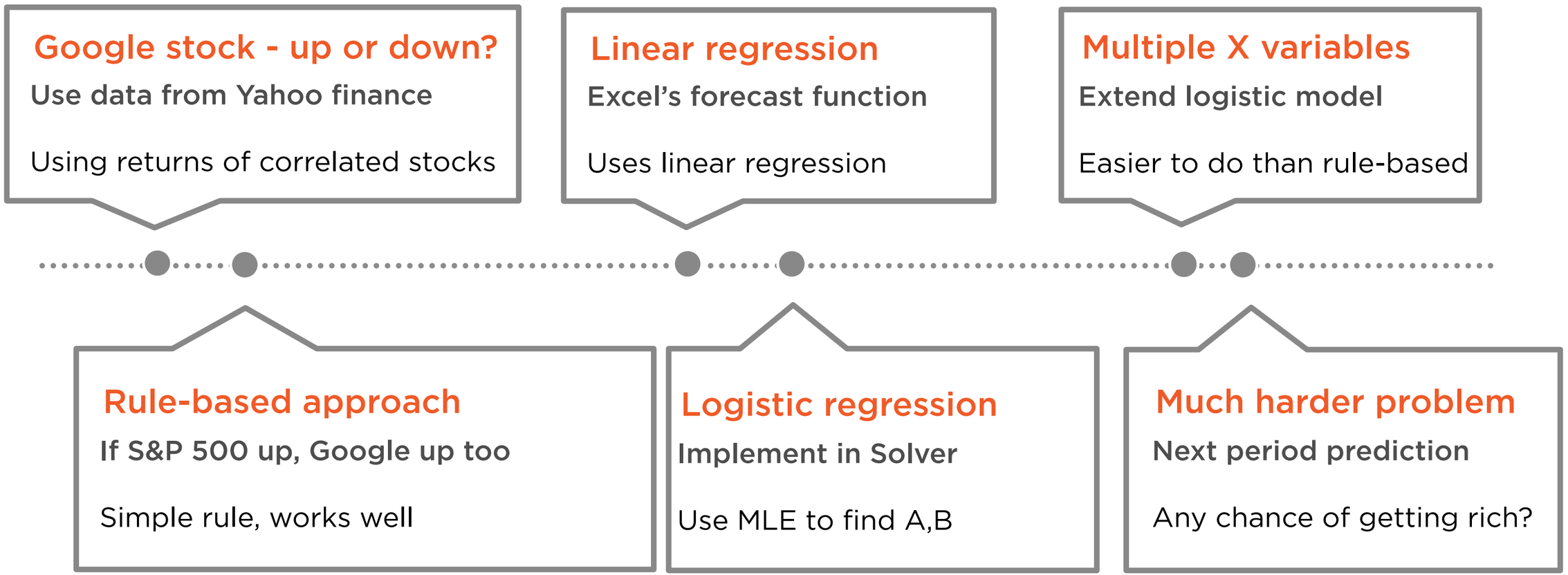
$$p(y_i) = \frac{1}{1 + e^{-(A + B^{GOOG} x^{GOOG}_{i-1} + B^{SP500} x^{SP500}_{i})}}$$

P(y) = Probability of Google going up in the current month i

x^{GOOG} = Returns on GOOG for previous month

x^{SP500} = Returns on S&P 500 for current month

Logistic Regression in Excel



A Much Harder Problem

$$p(y_i) = \frac{1}{1 + e^{-(A + B^{GOOG} x^{GOOG}_{i-1} + B^{SP500} x^{SP500}_{i-1})}}$$

P(y) = Probability of Google going up in the current month i

x^{GOOG}_{i-1} = Returns on GOOG for previous month

x^{SP500}_{i-1} = Returns on S&P 500 for previous month

Very difficult problem to solve - quant hedge funds are very interested in the answer

Summary

Logistic regression can be easily implemented in Excel using Solver

Applying this to explaining stock returns yields similar results to a rule-based approach

Multiple explanatory variables are far easier to add in the logistic approach