Implementing Our Regression Solution

Mohammed Osman
SENIOR SOFTWARE DEVELOPER
@cognitiveosman www.cognitiveosman.com

Overview

Pipeline once again

After all: What is Model Training?

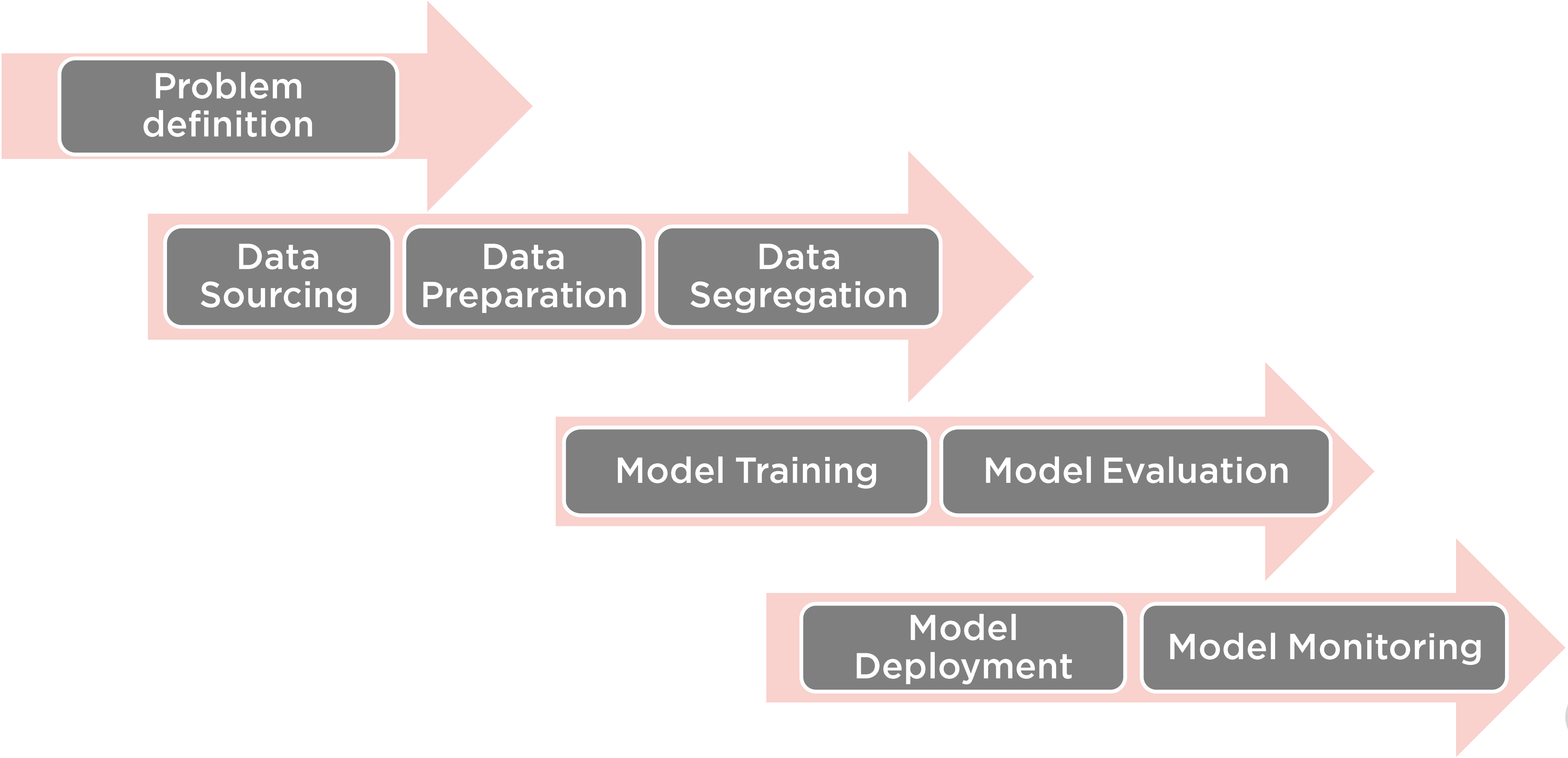
Foundational Concepts

Linear Regression

Model Evaluation

Demo

Model Training and Evaluation



What Is Model Training?

All Machine Learning algorithms use one principle

Three types of Machine Learning algorithms (generally)

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Our focus is Supervised Learning

Types of Supervised Learning Algorithms

Regression

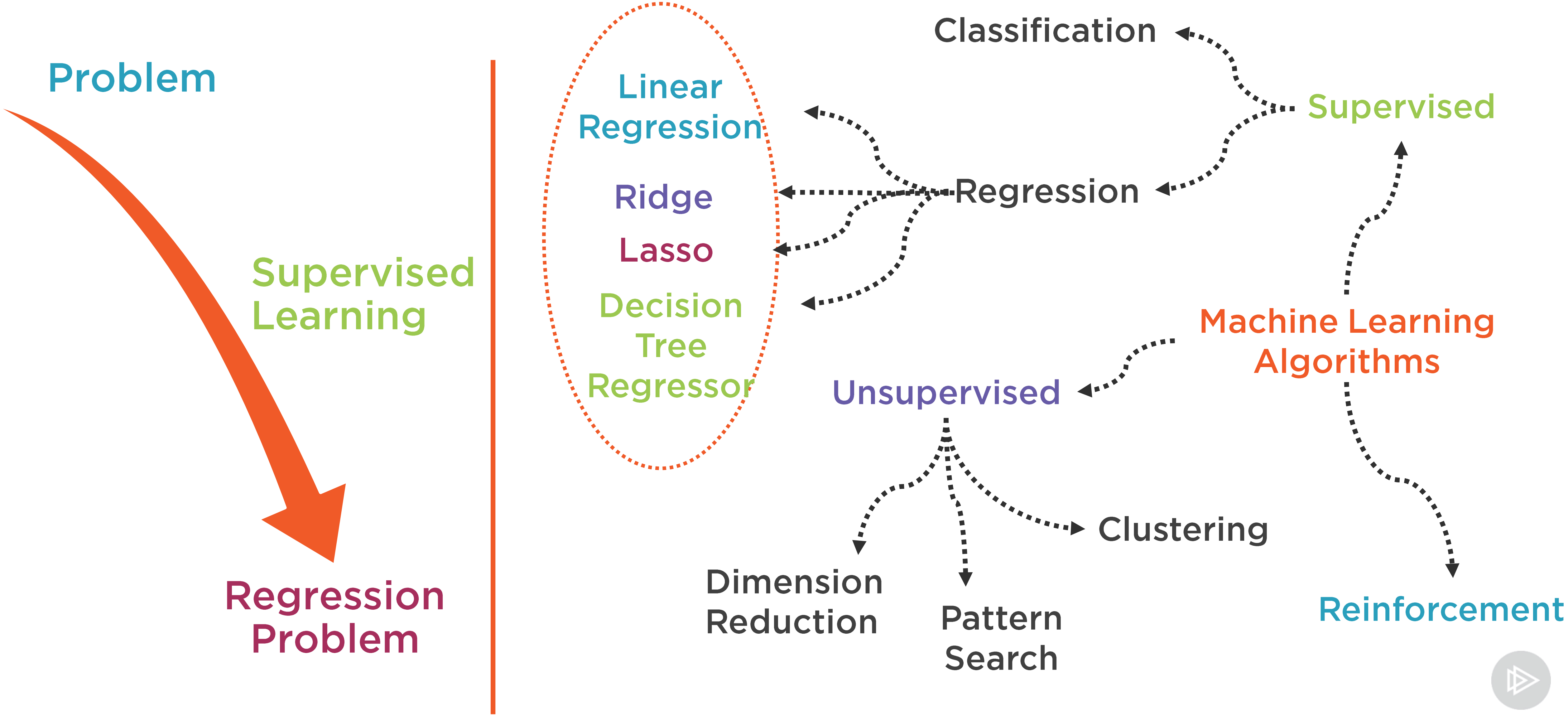
For continuous values

Classification

For discrete (categorical) values

In the context of Supervised Learning:
Machine Learning training is the process of learning an ML **algorithm** how to find patterns in the input data so that they correspond to the target, resulting a machine learning model

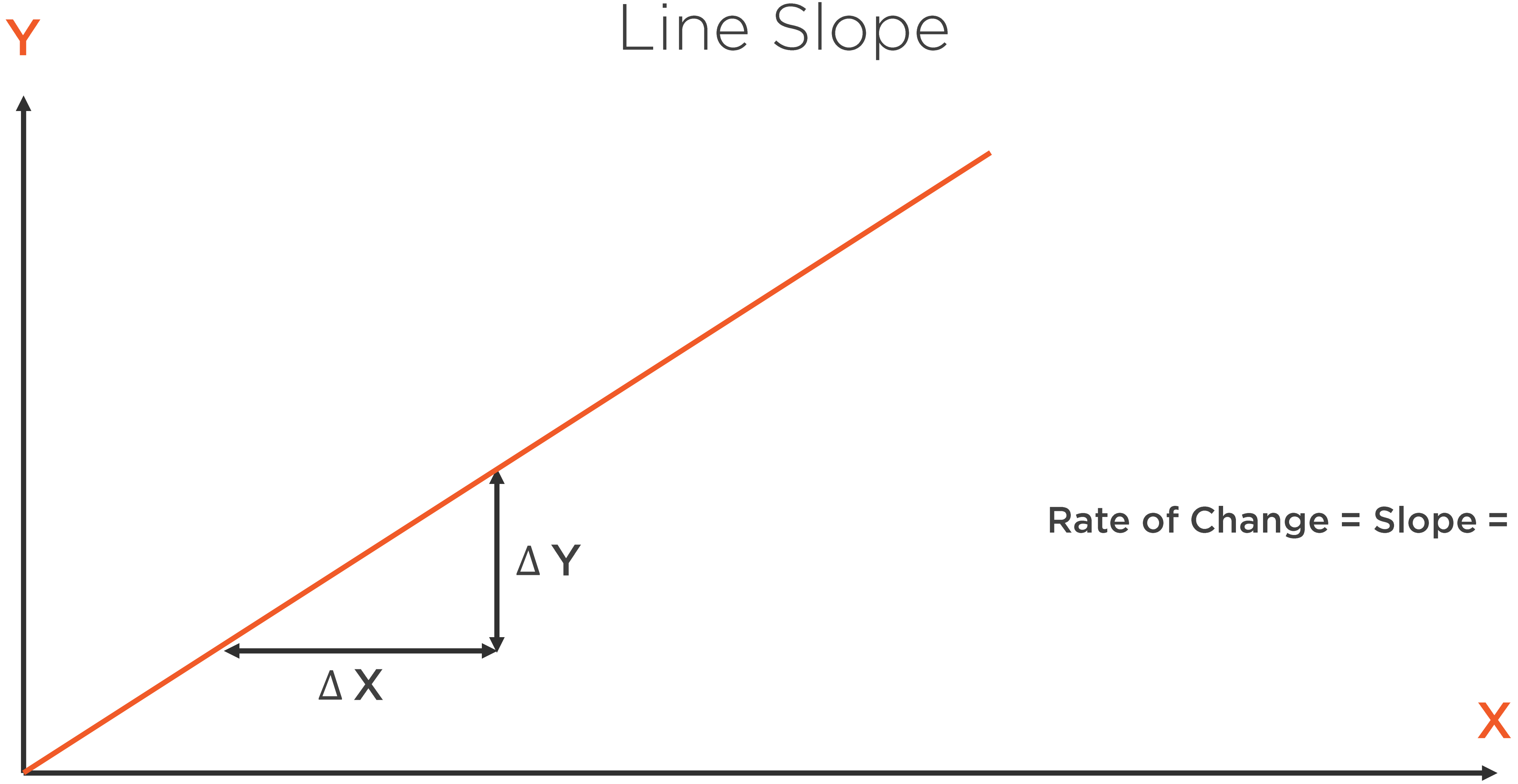
Why Should I Care?



Learning More

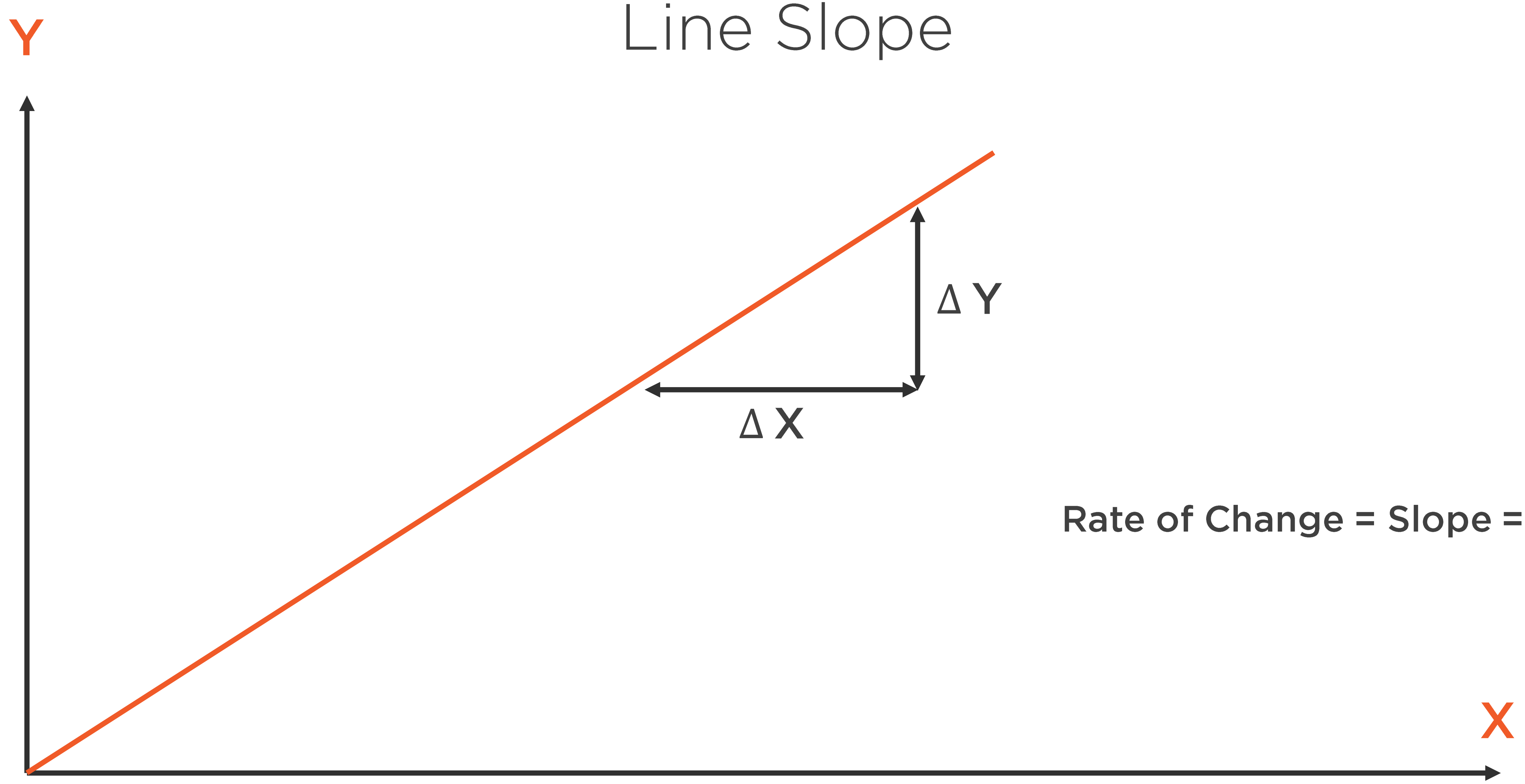
Foundational Concepts

Line Slope



$$\text{Rate of Change} = \text{Slope} = \frac{\Delta y}{\Delta x}$$

Line Slope



Rate of Change = Slope = $\frac{\Delta y}{\Delta x}$

Important Line Slope Cases

Horizontal Line

Slope = 0

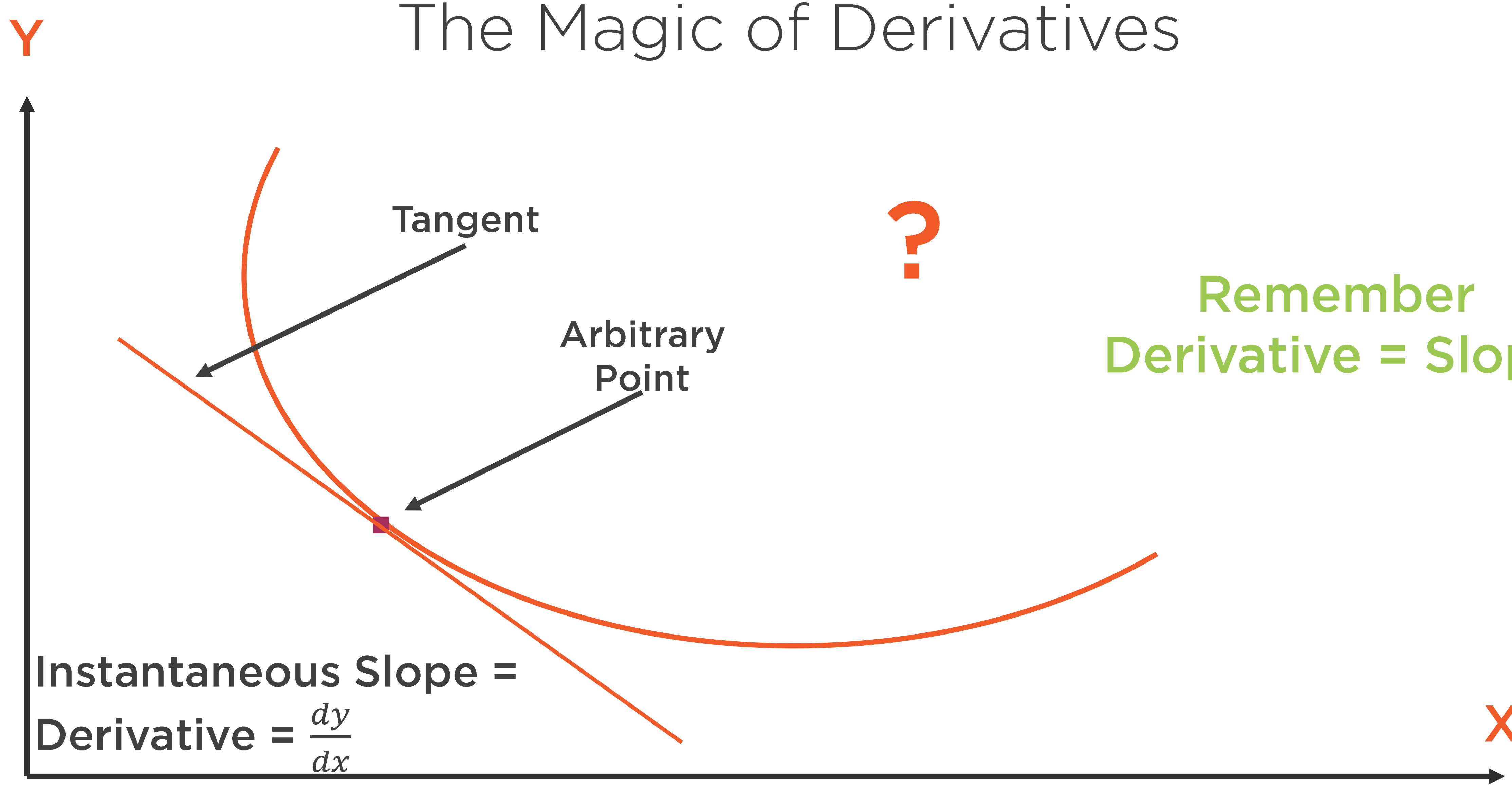
Vertical Line

Slope = Infinity

45 Degrees Line

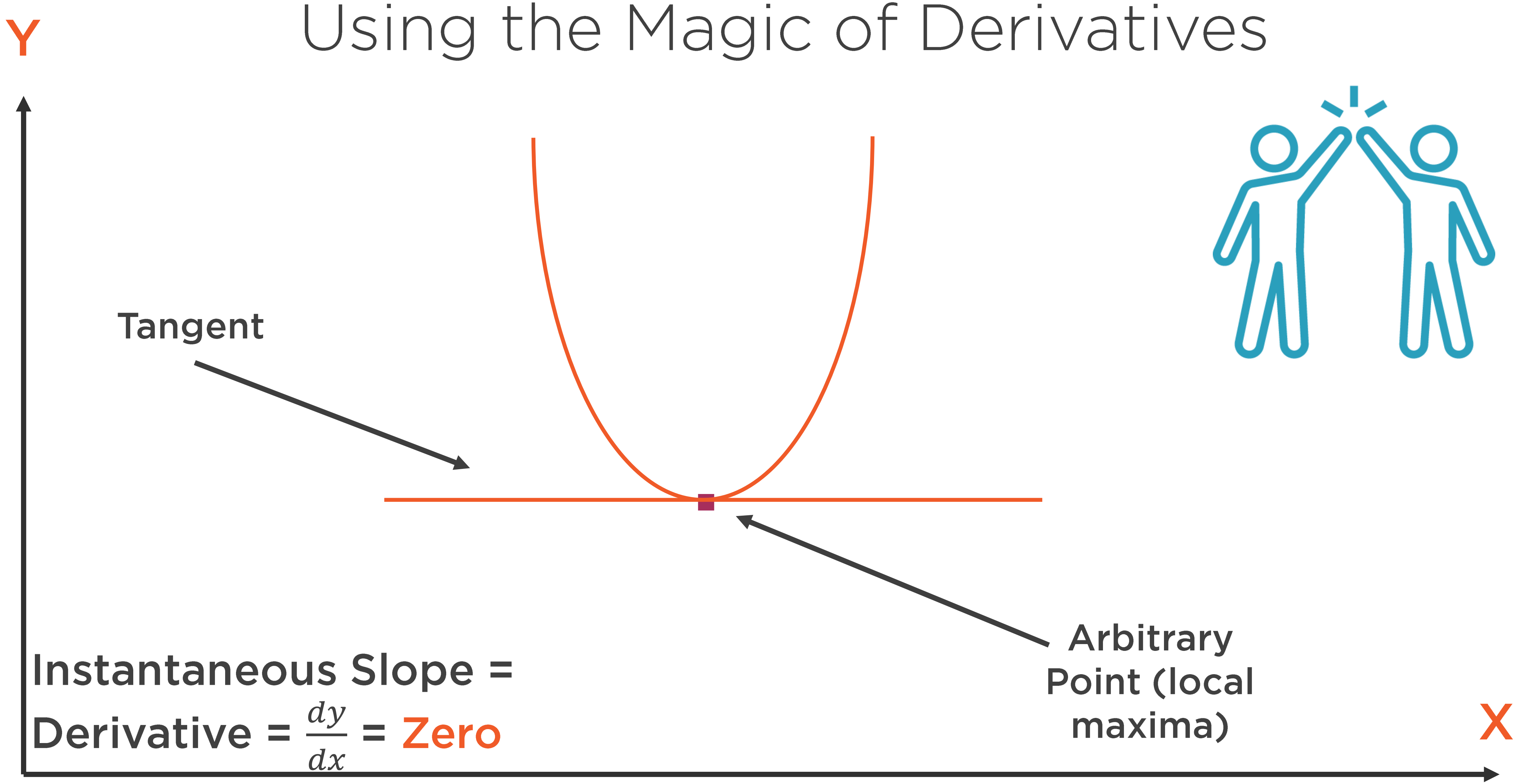
Slope = 1

The Magic of Derivatives



Remember
Derivative = Slope

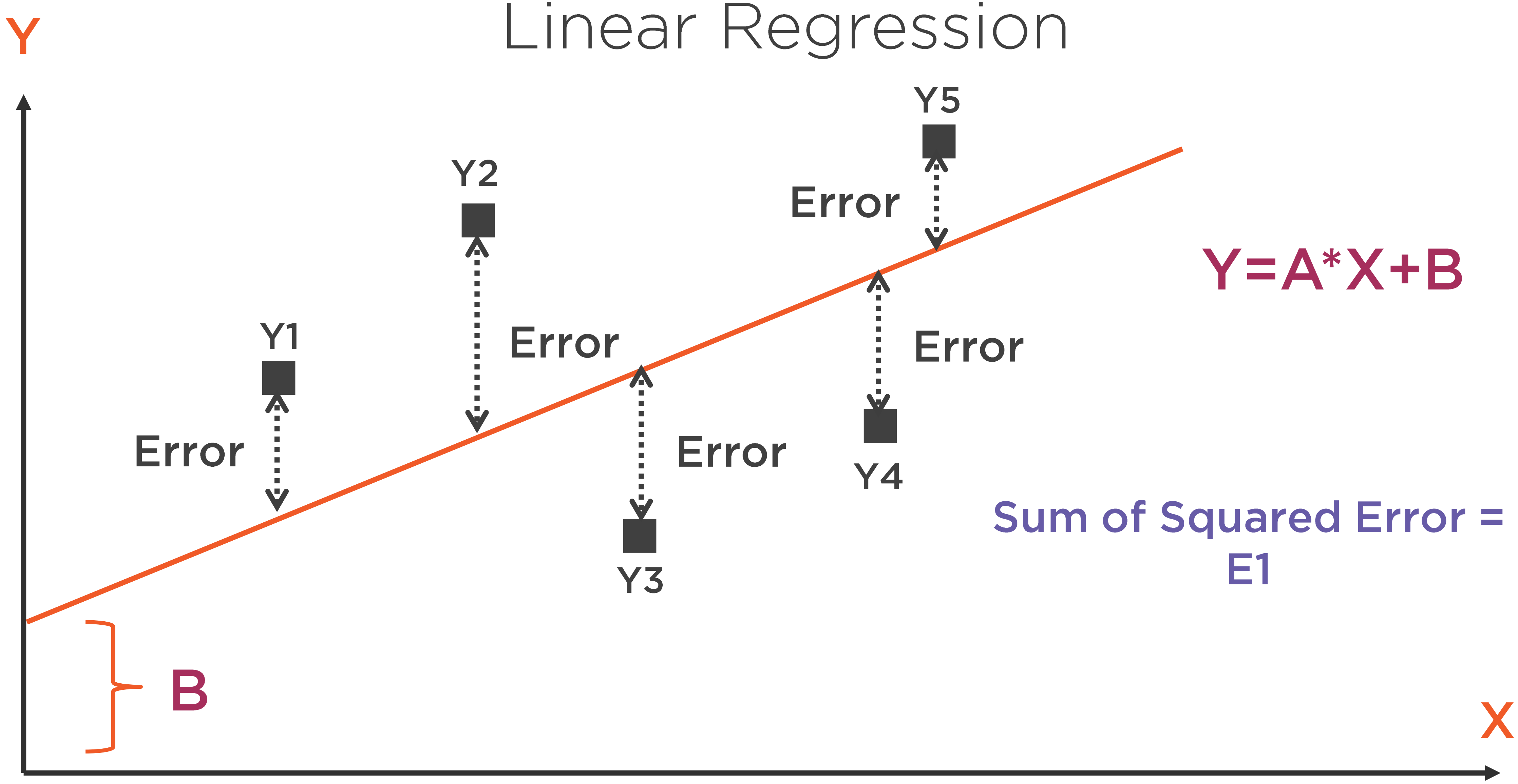
Using the Magic of Derivatives



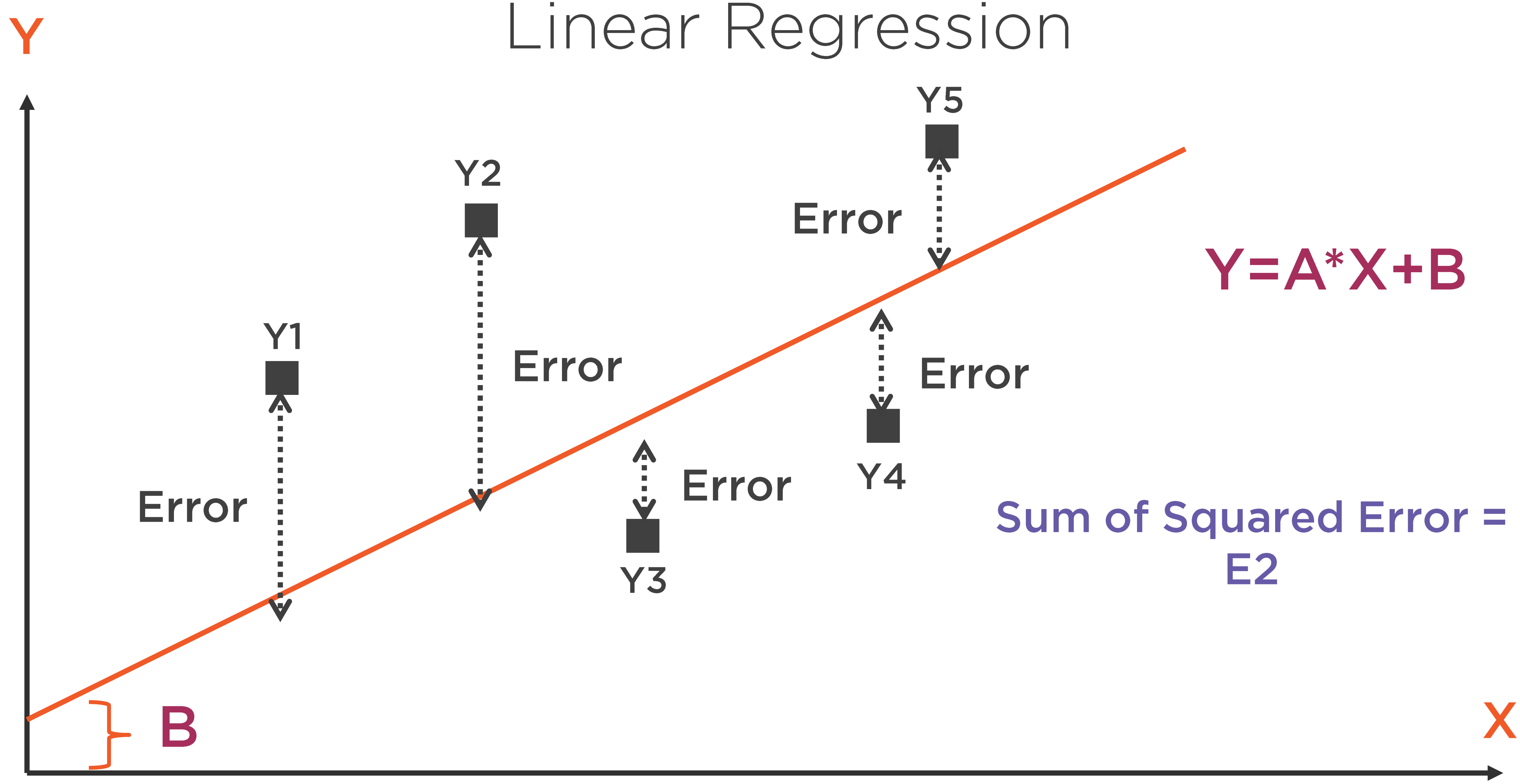
If we want to get the minimum of a function, then we calculate its derivative when it is equal to zero!

Linear Regression Algorithms

Linear Regression



Linear Regression



How Linear Regression Works?

The of sum of residuals is calculated as $\sum(Yn - (AXn + B))^2$

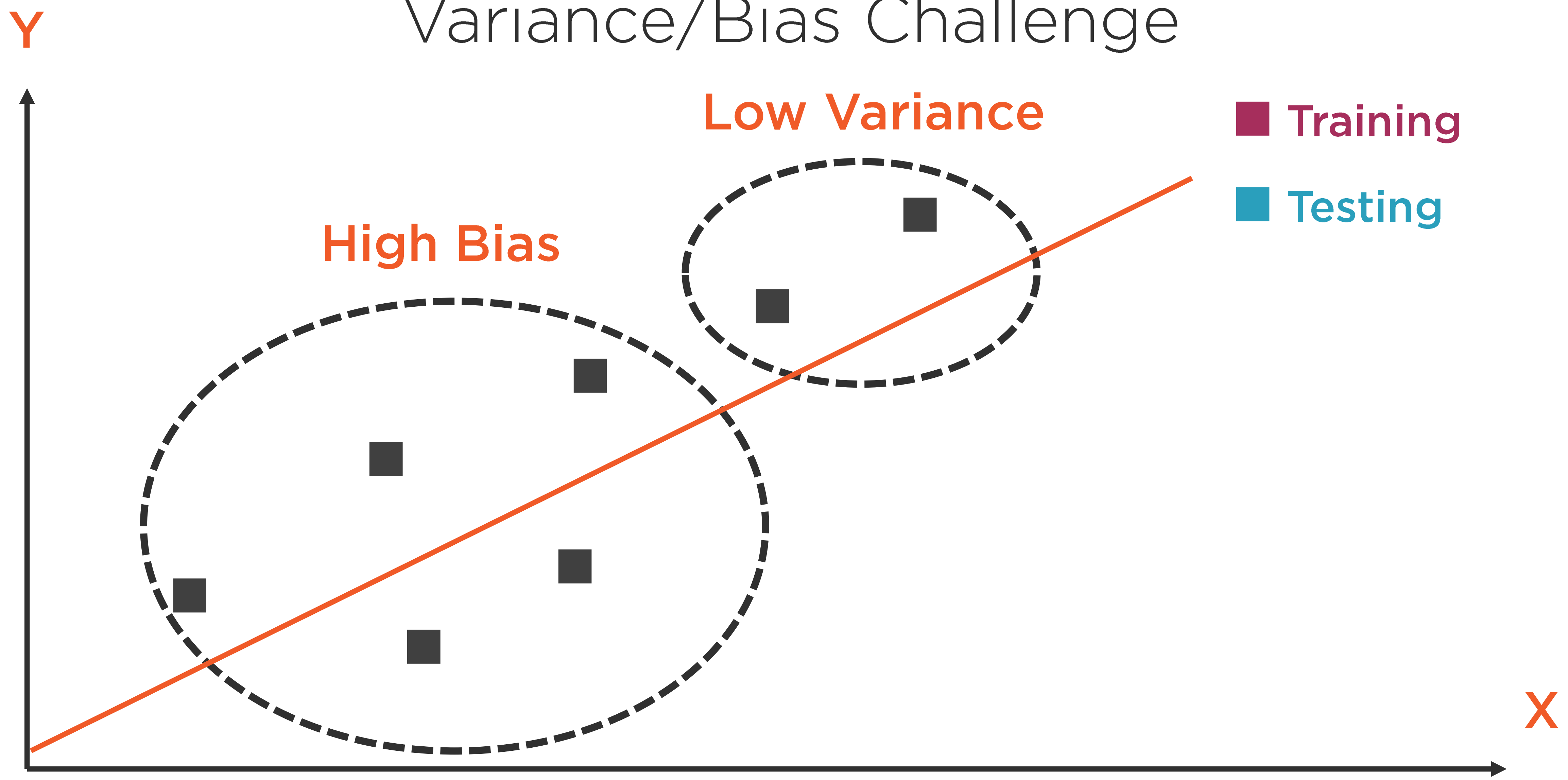
The sum of residuals minimized by taking the derivative at zero with respect to the slope

The sum of residuals minimized by taking the derivative at zero with respect to the intercept

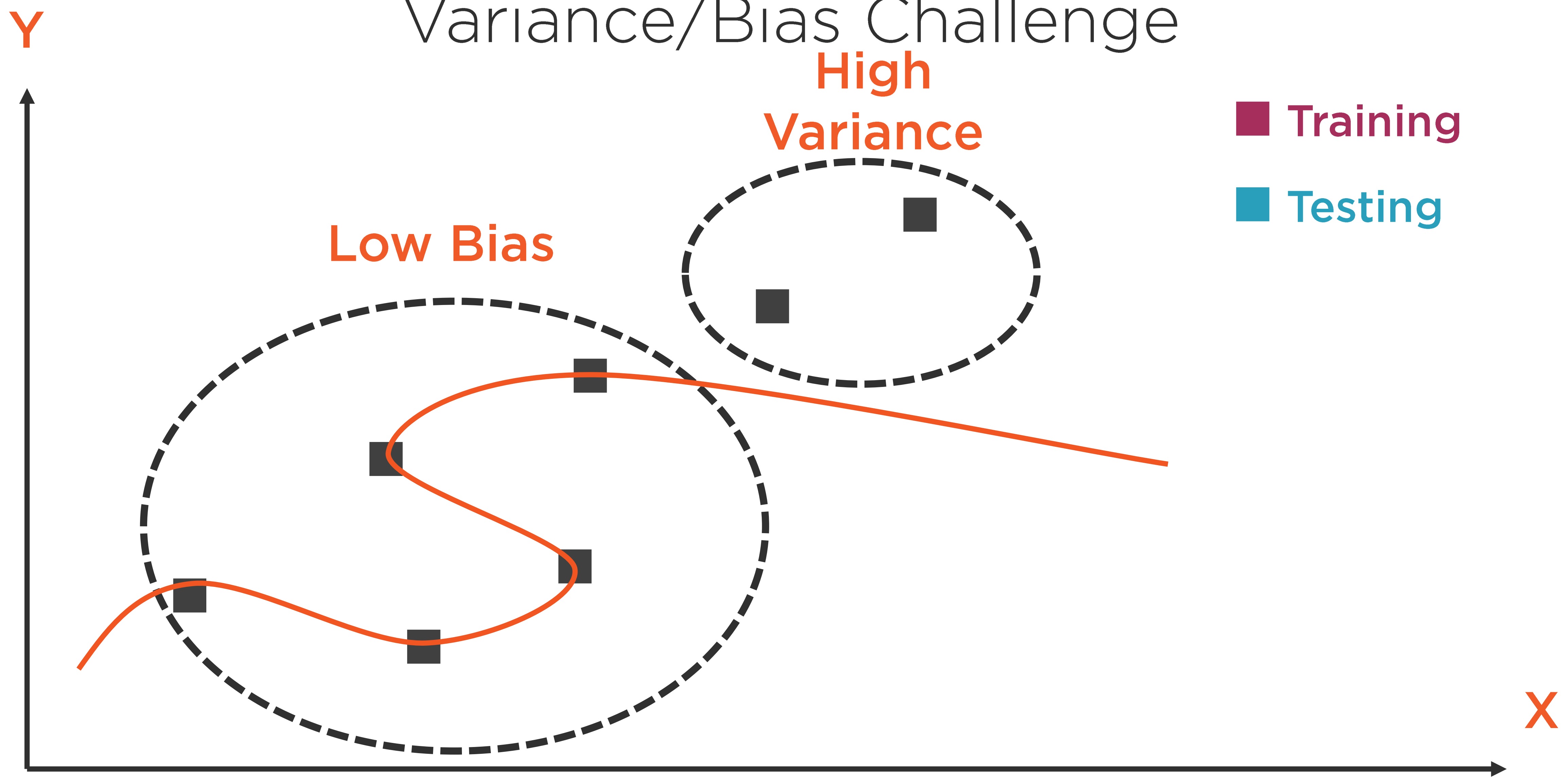
We solve the equations to get the slope (A) and intercept (B)

We solve for the the slope (A) and the intercept (B)

Variance/Bias Challenge



Variance/Bias Challenge



Variance/Bias trade-off

Variance

$$Var(x) = \sigma^2 = \frac{\sum(\bar{x} - u)^2}{N}$$

Occurs when the ML algorithm has high ability to fit training data

Error due to fluctuations in the training set

High variance usually indicates overfitting

Bias

Occurs when a ML algorithms has limited ability to learn

Wrong assumption about the problem nature

High bias usually associated with underfitting

Regularization

Is the process of tuning model parameters or complexity so that the model performs better at predicting (generalizing) on out of sample data.

Linear Regression Regularization

Ridge Regression

Be concise and keep the text to four lines or fewer

Lasso Regression

Be concise and keep the text to four lines or fewer

Elastic Net Regression

Combines techniques from Ridge and Lasso

Other Regression Algorithms

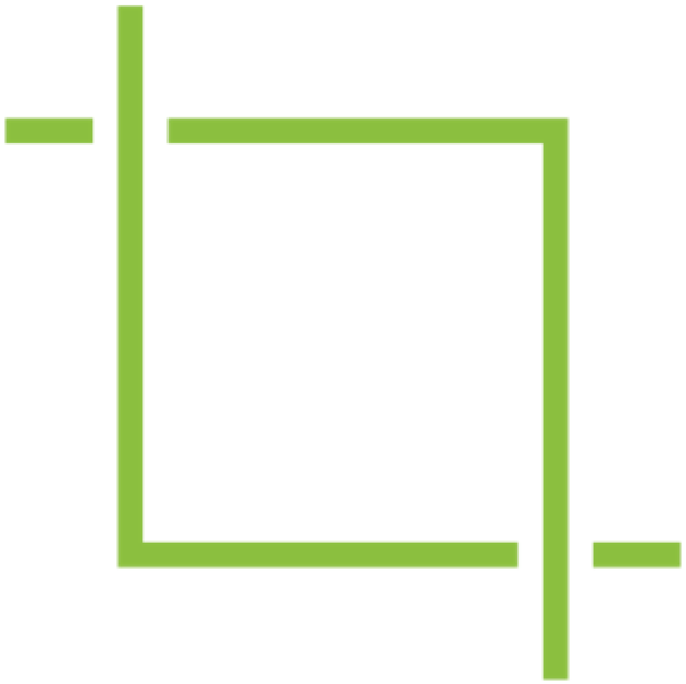
K-neighbors Regression

Simple algorithm

Relies on distance measurement

Requires standardization

Support Vector Regression (SVR)



Used for both regression and classification (SVM)

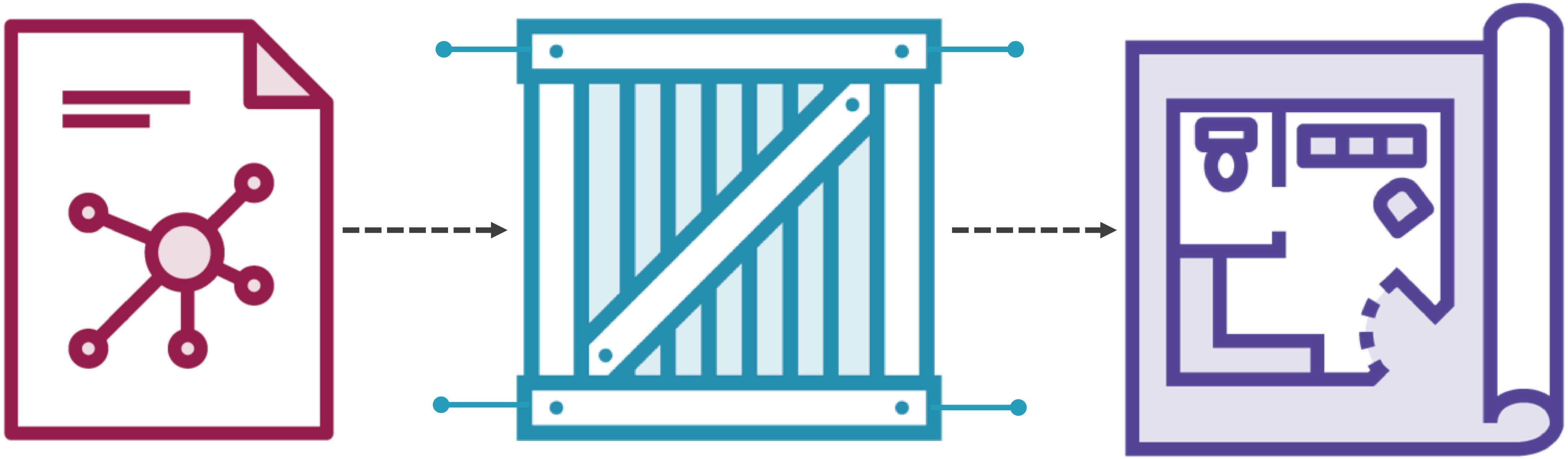
Similar to linear regression

Decision Tree Regressor

Used for both regression and classification

Reaches the answer by structuring the data in tree leaves

Should I Care?



Data ML algorithm ML model

Model Evaluation

Regression Models Evaluation Metrics

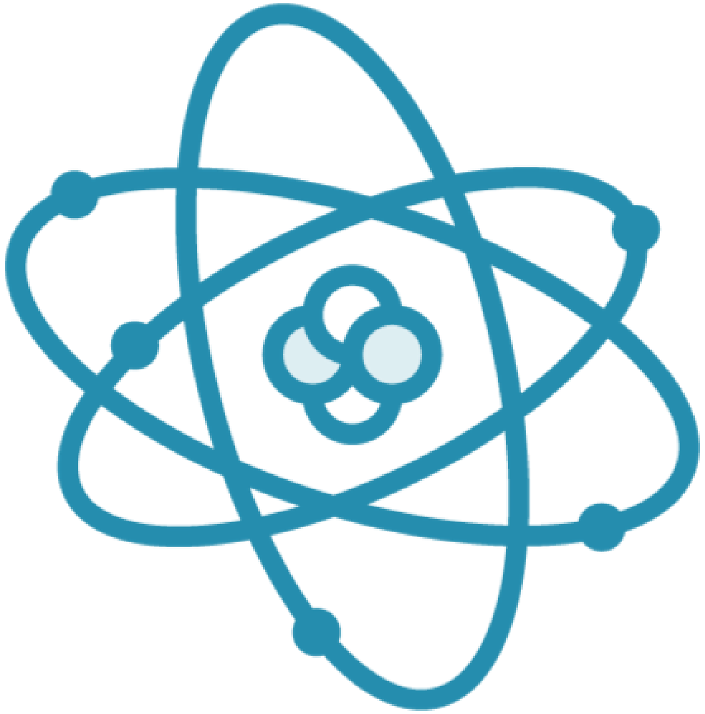
- Max Error
- Mean Absolute Error
- Mean Squared Error
- R Squared
- Others

Max Error

Captures the worst case

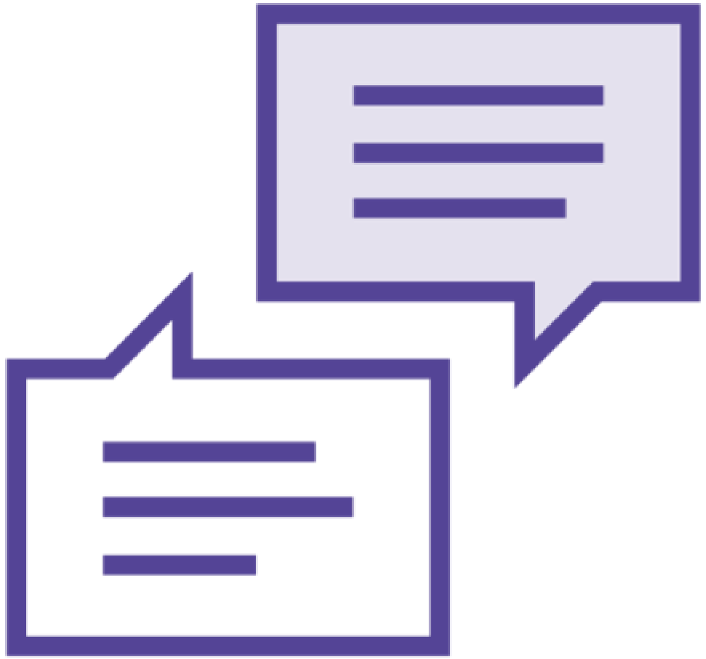
How much we can tolerate

Mean Absolute Error



Average of absolute errors

Mean Squared Error



Average of squared errors

MAE vs. MSE

Mean Absolute Error

Removes negative signs by taking absolute value

More robust to outliers

Mean Squared Error

Removes negative signs by squaring the values

Better when we want to penalize outliers

R^2 (Coefficient of Determination)

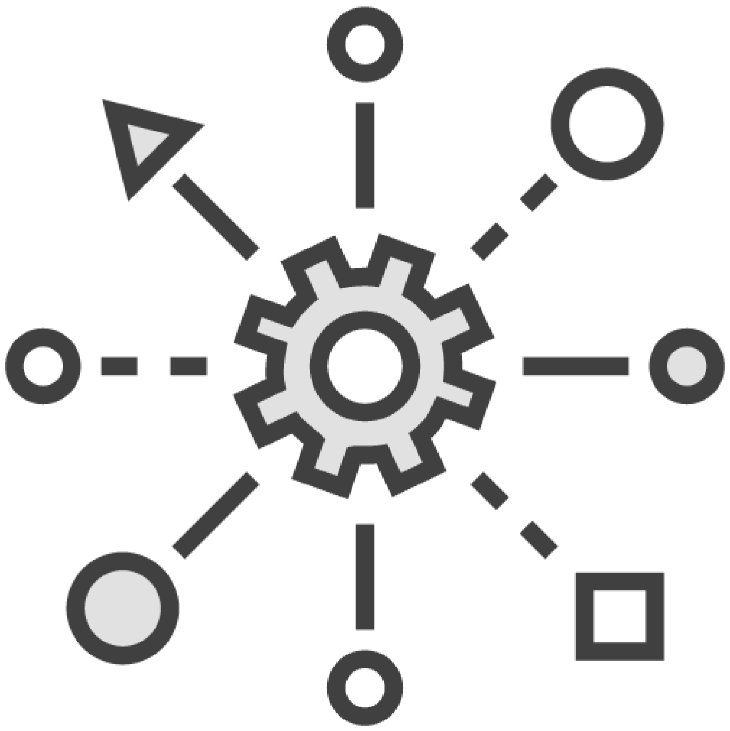
$$R^2 = \frac{Var(mean) - Var(fit)}{Var(mean)} = Correlation^2$$

Tells us how much percentage of the data is explained by the relationship but no direction

Var(mean) = 22 and Var(fit) = 4 then R^2=0.81

Easier to interpret

Others



Median absolute error

Mean squared log error

Demo

Data Segregation

- Train/Test split
- K-Fold Cross Validation

Glitch with Scikit-learn

Summary

Final look to ML pipeline

What is model training?

Foundational concepts

- Slope
- Derivative

Linear regression algorithms

Variance/Bias trade-off

Other regression algorithms

Model evaluation

Demo