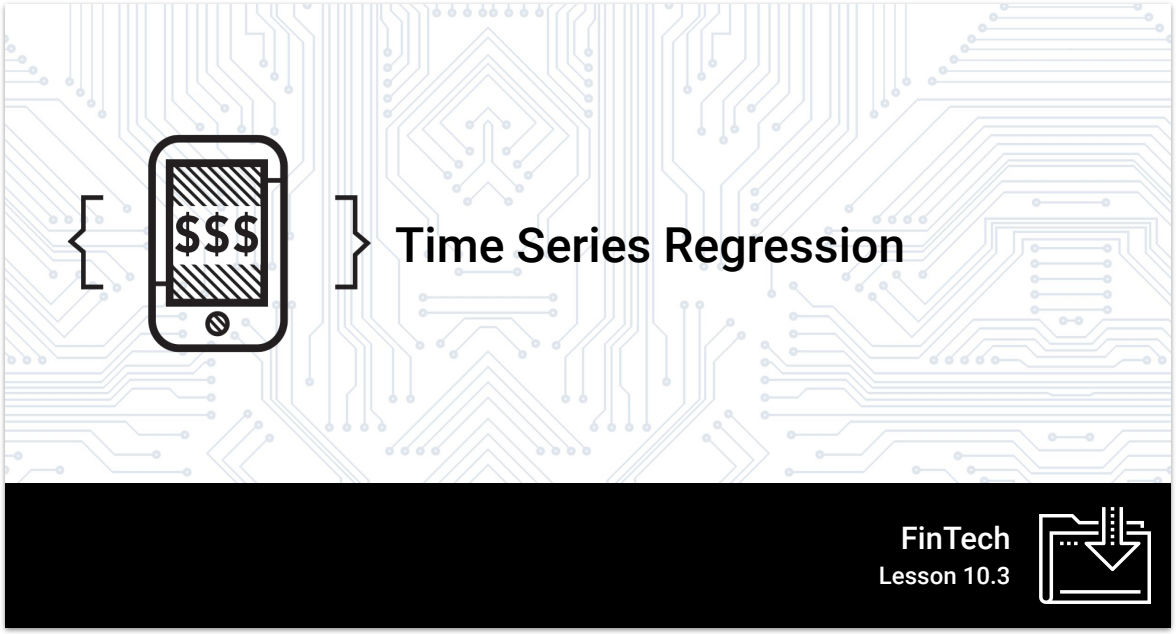


Class Objectives

By the end of today's class you will understand:

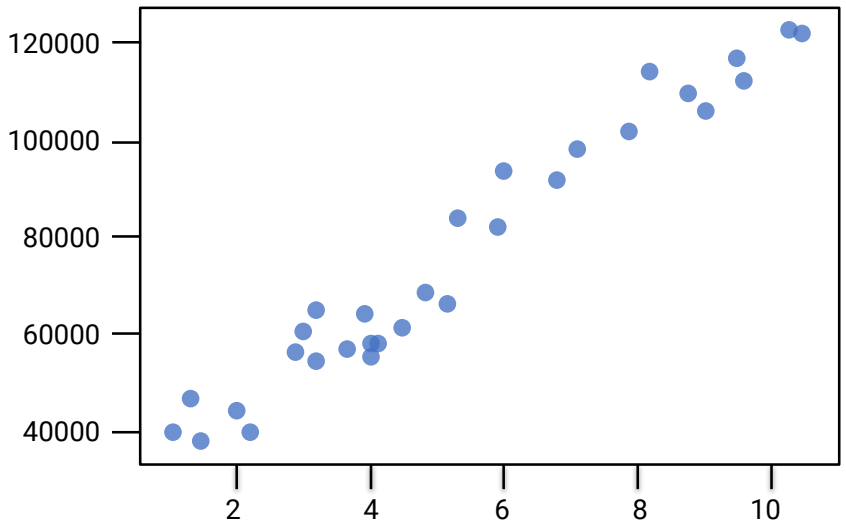
- Linear Regression
- Time Series Linear Regression
- Regression Metrics
- Train Test Split
- Rolling Out-of-Sample

Linear Regression

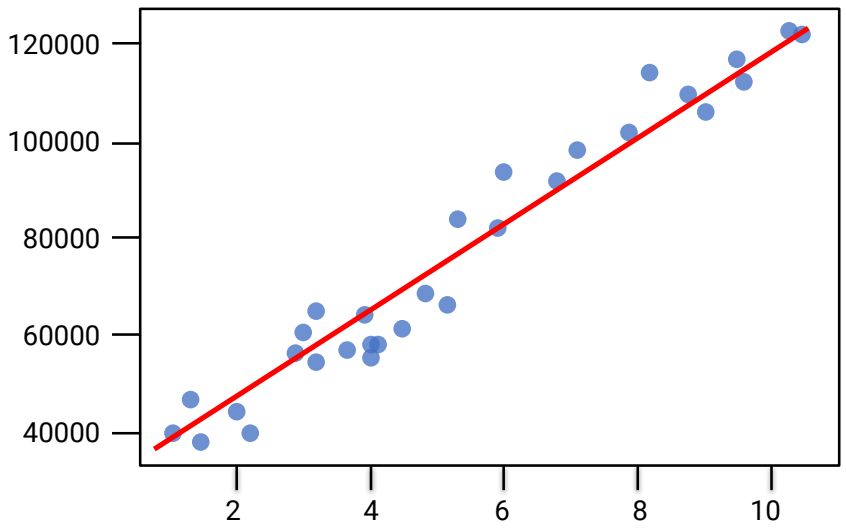
Line Equation

y	=	mx + b
m	=	slope
b	=	y-intercept (the value of y when x = 0)

Linear Regression: Find the Line That Best Describes the Data



Best Fit Line



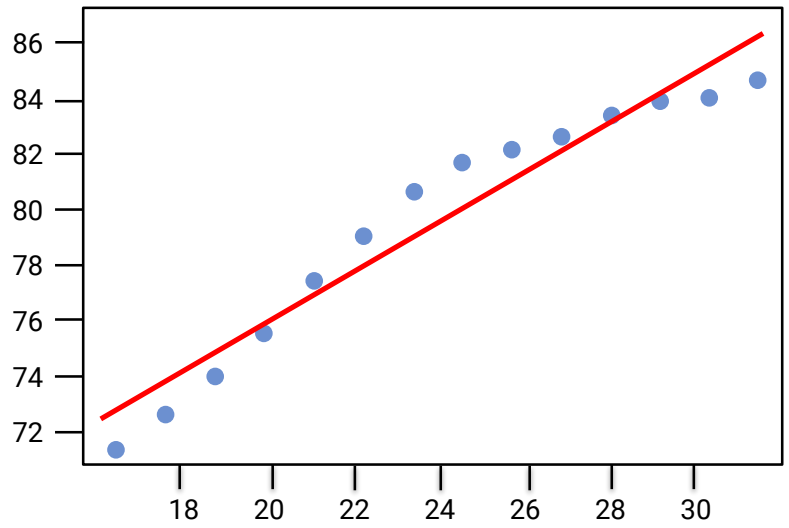
Multiple Regression

Each day (X) is assigned its weight, or coefficient.

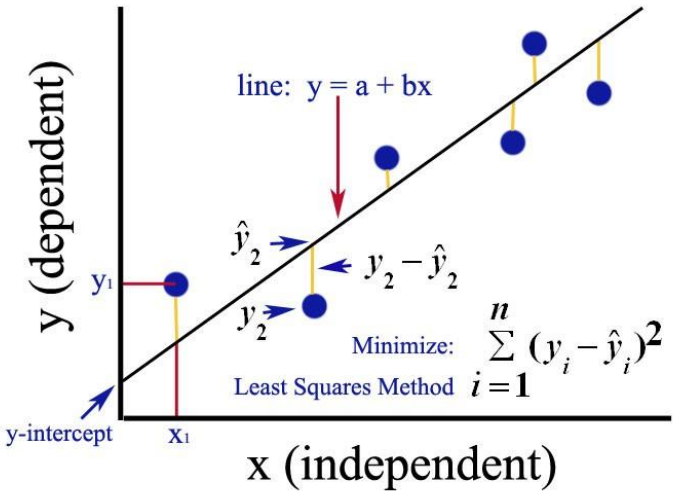
$$y = b_0 + b_1 X_1 + b_2 X_2 ...$$

Regression Metrics

Best Fit Line



Regression Metrics

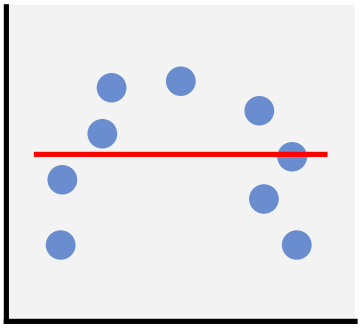


AdobeStock_56885471

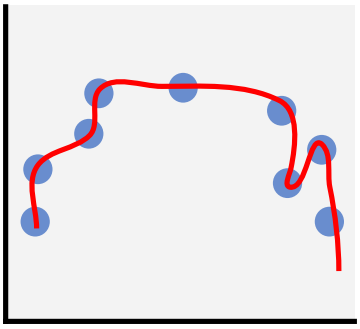
Overfitting

Overfitting

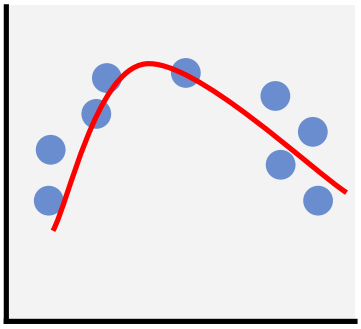
Underfit



Overfit

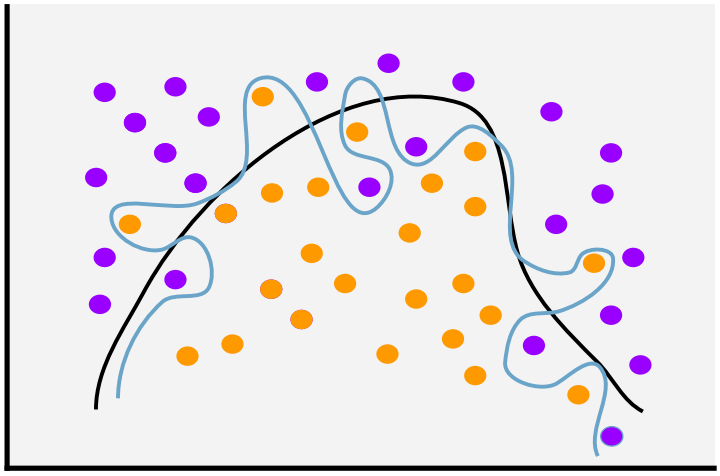


Ideal

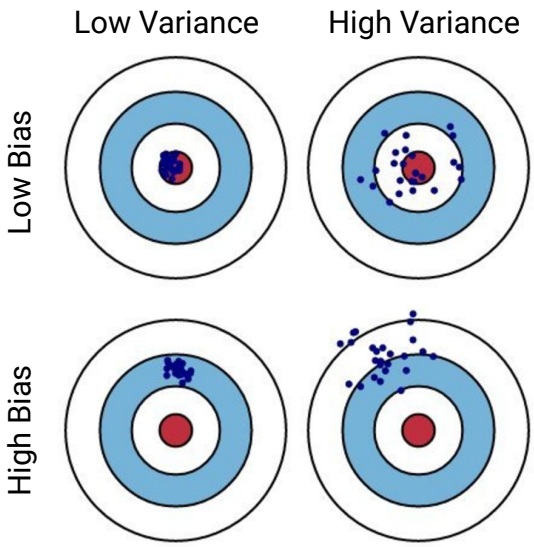


Overfitting

Overfit models learn the ‘noise’ found in the training data, rather than just the ‘signal’



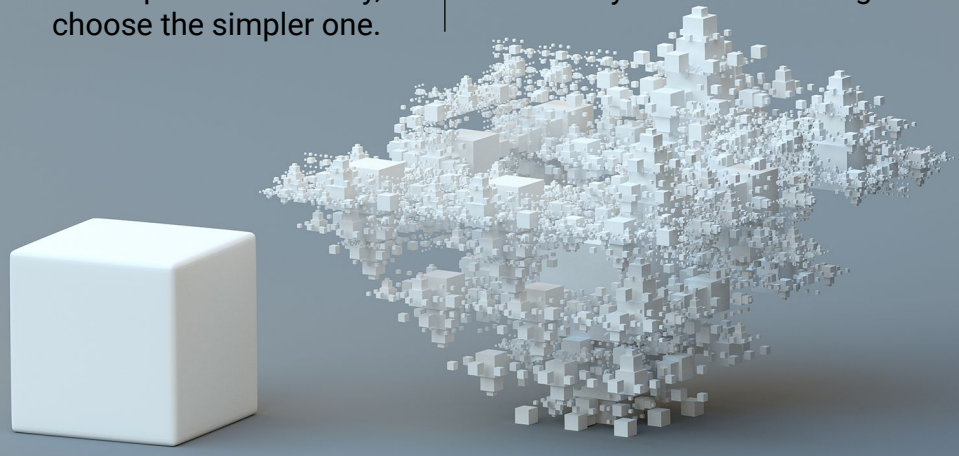
Variance vs Bias



Parsimony

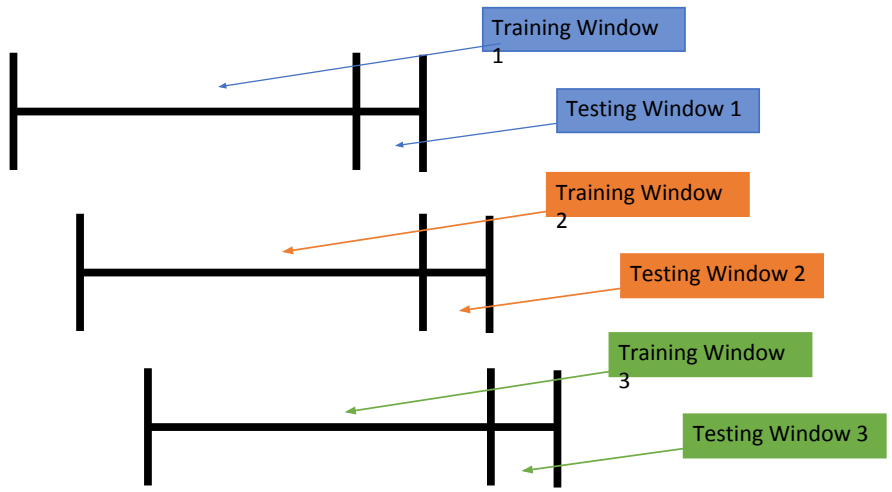
Statistical application of Occam's razor: when two models perform similarly, choose the simpler one.

Why? Needlessly complex models are harder to compute and may lead to overfitting.



Rolling Out-of-Sample

A Rolling Out-of-Sample Approach



Questions?