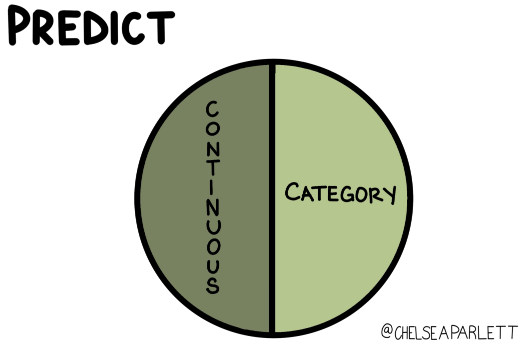


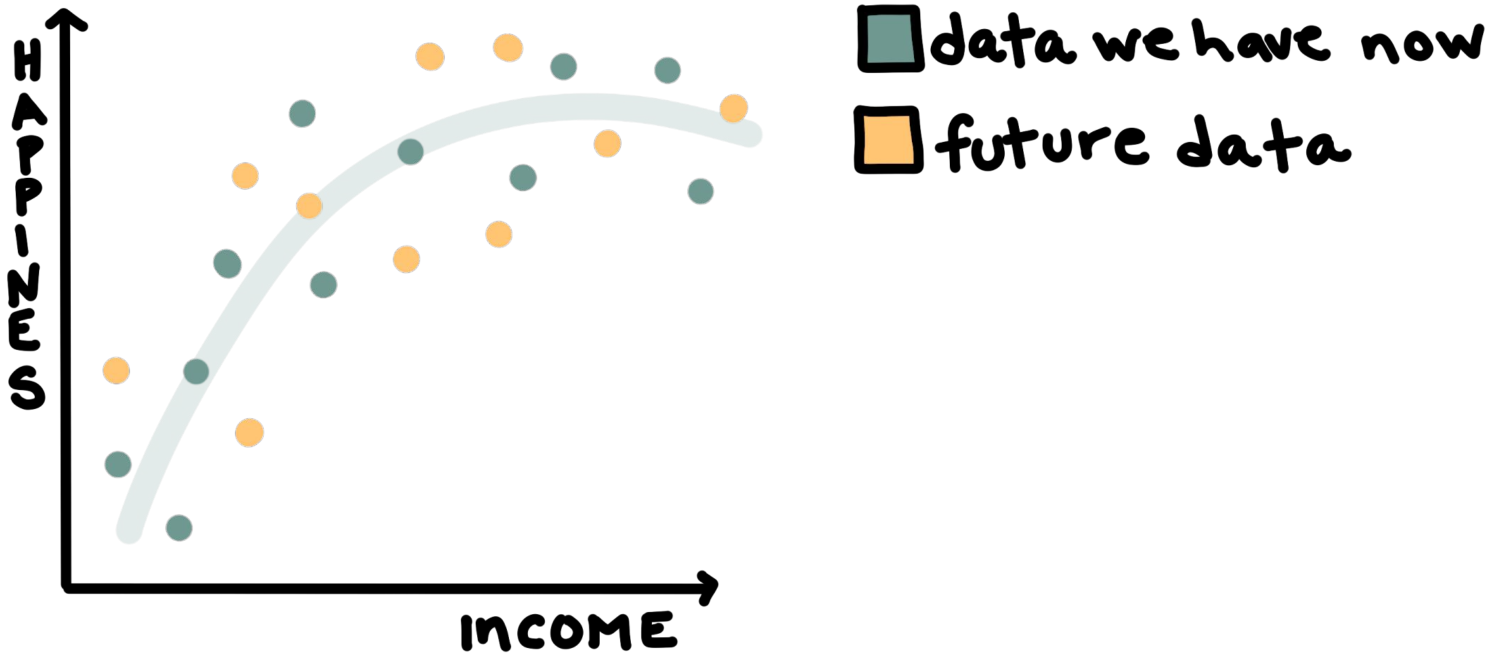
Linear Regression III

Dr. Chelsea Parlett-Pelleriti

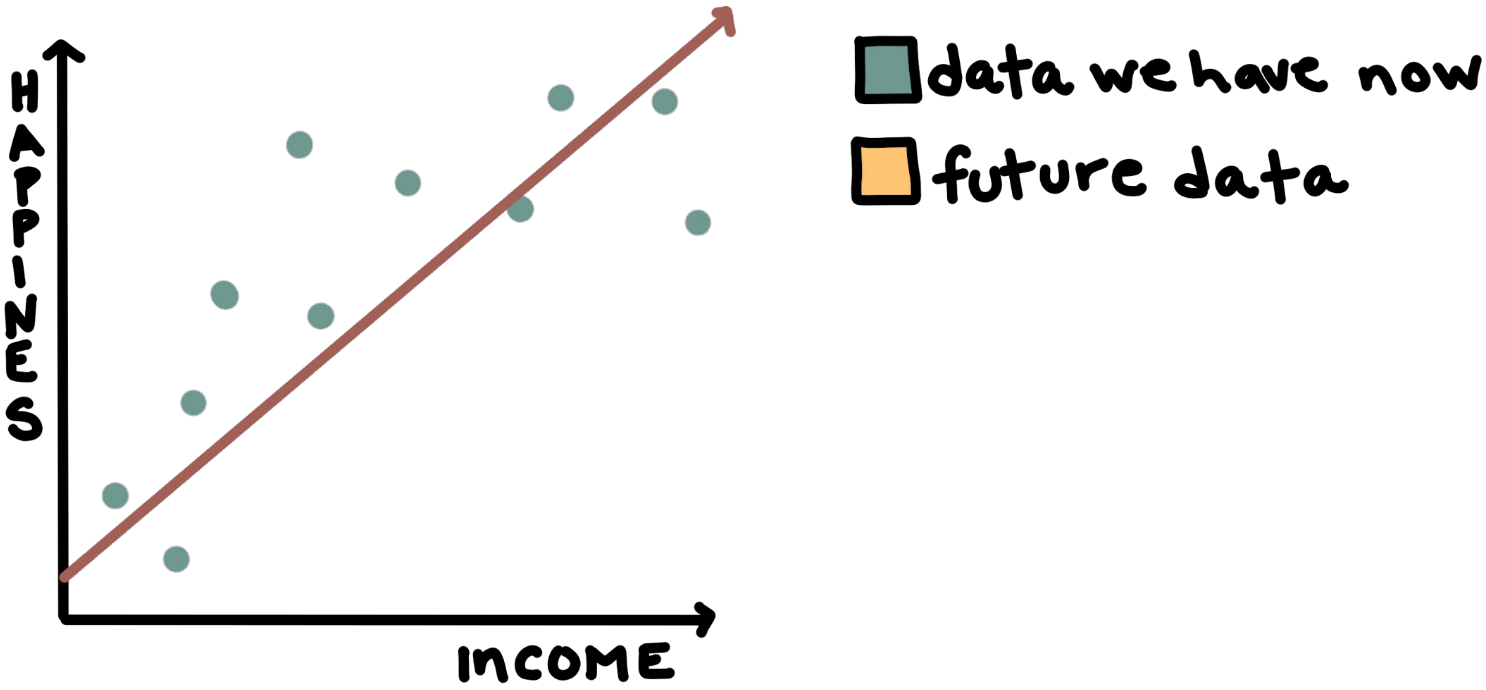
Linear Regression

- Bias and Variance Definition
- Bias Variance Tradeoff
- Model Validation
- Double Descent

Bias and Variance



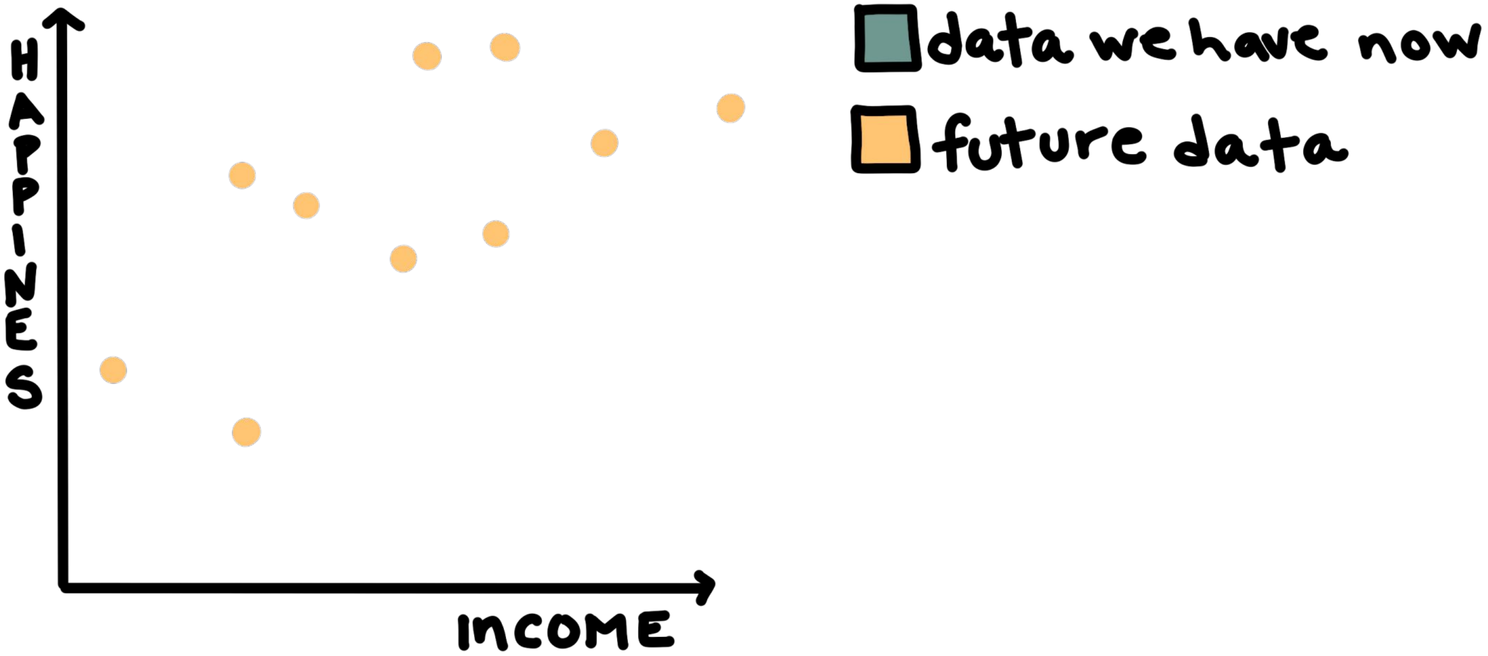
Bias and Variance



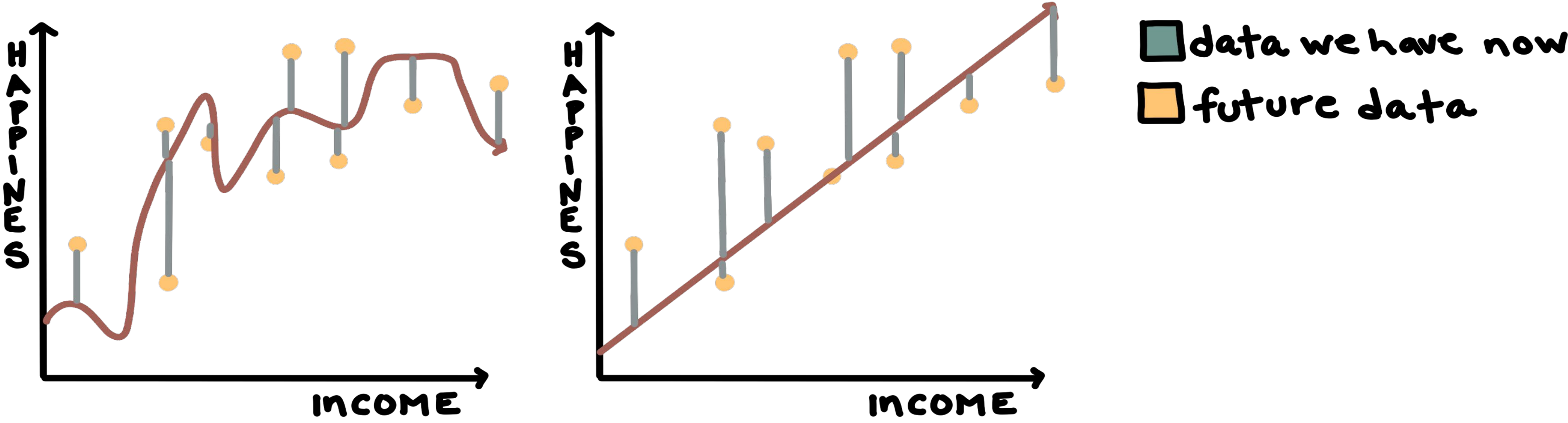
Bias and Variance

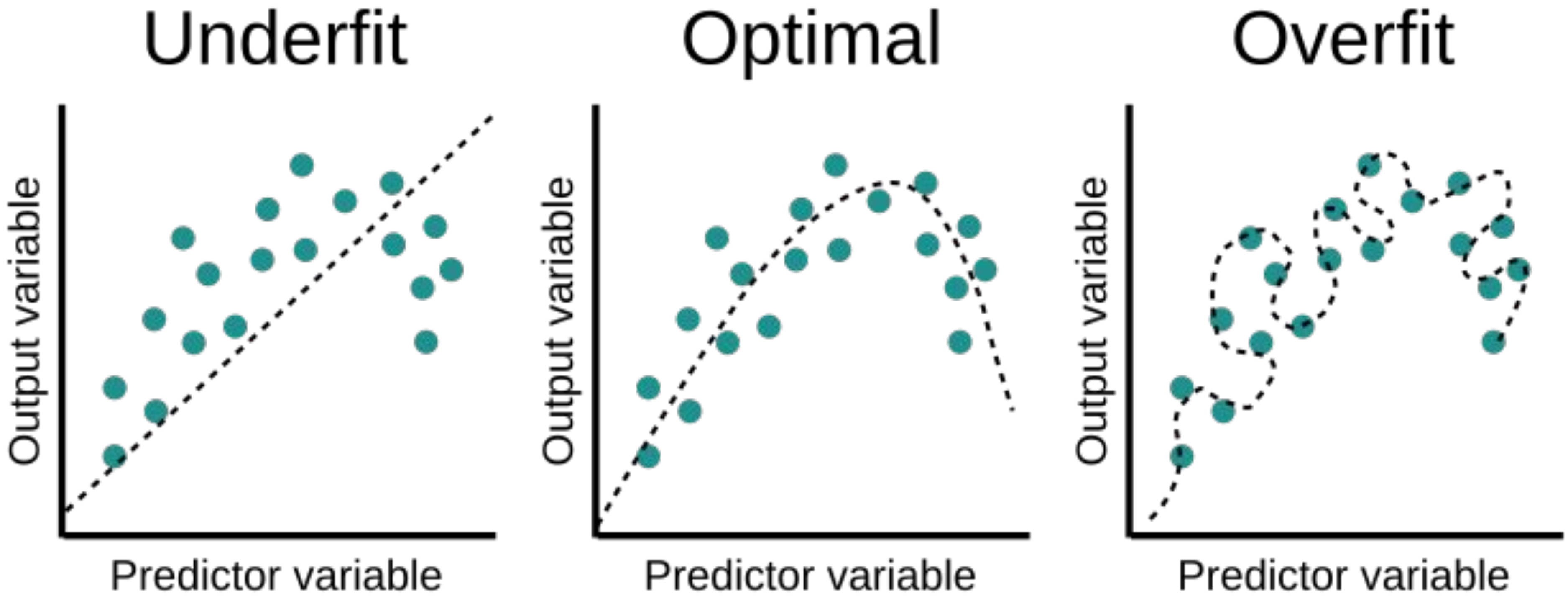
Bias and Variance

Bias and Variance

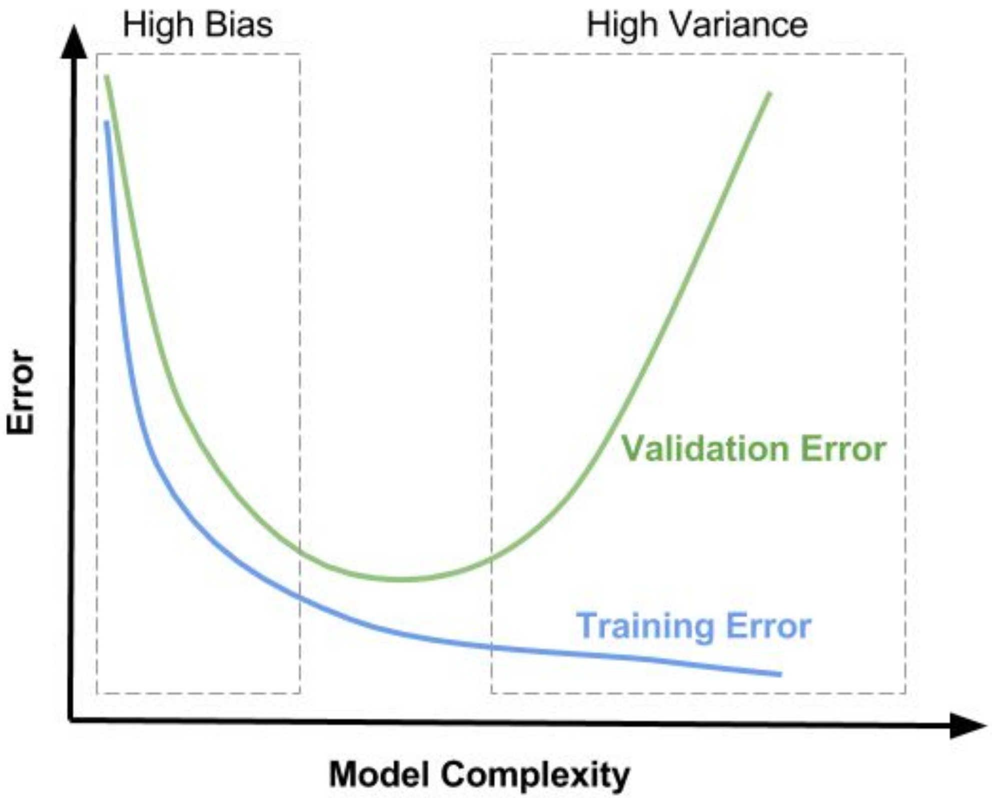


Bias and Variance



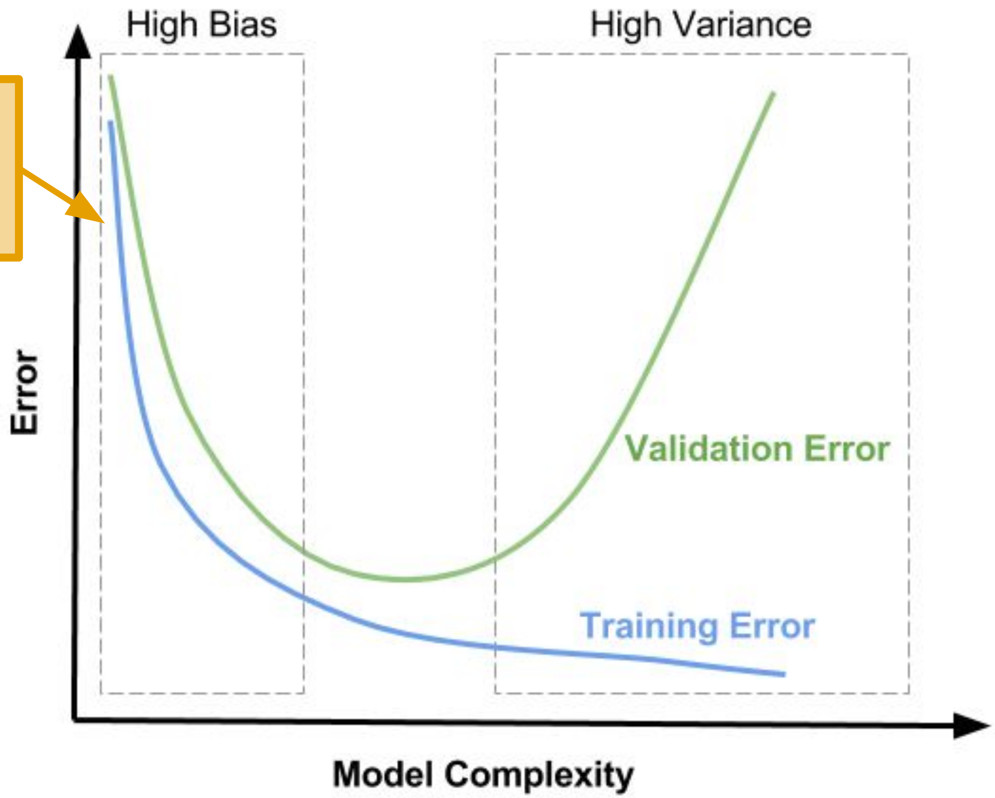

Underfit
Optimal
Overfit
Output variable
Predictor variable
Output variable
Predictor variable
Output variable
Predictor variable

Bias and Variance

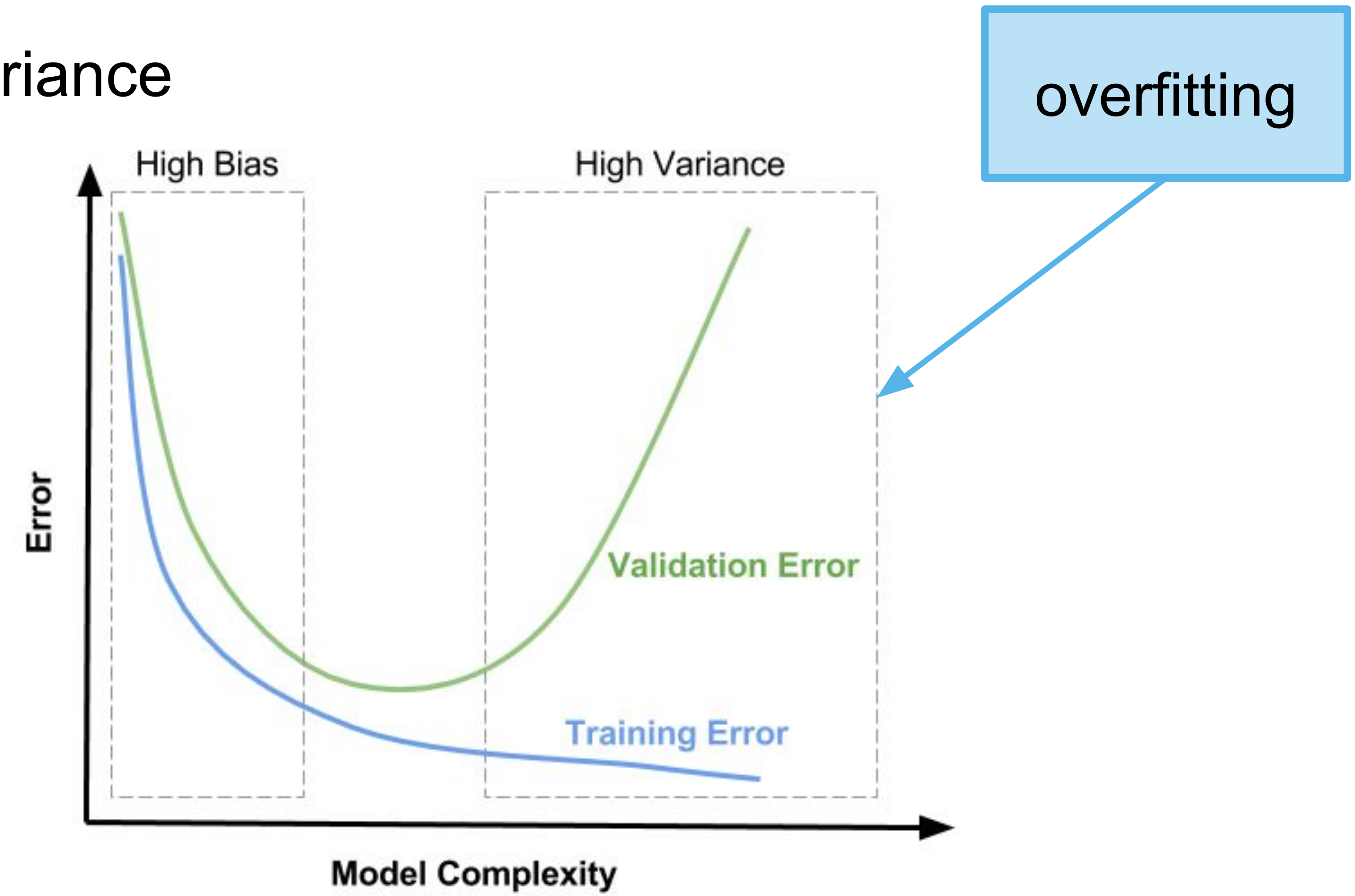


Bias and Variance

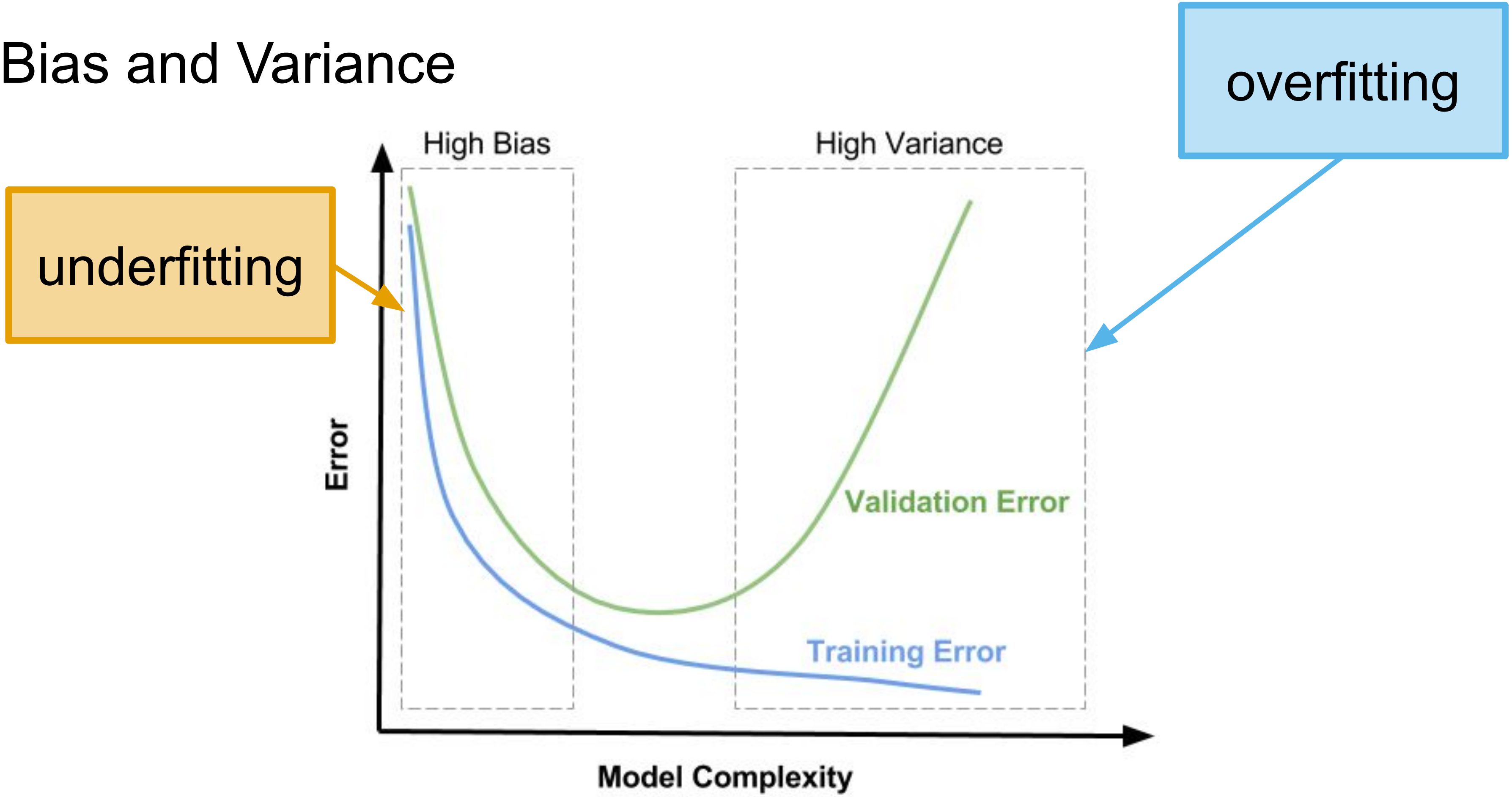
underfitting



Bias and Variance



Bias and Variance



Math

Our Goal

$$Y = f(X) + \epsilon$$

Our Goal

$$Y = f(X) + \epsilon$$

$$\hat{f}(X)$$

Sources of Error

$$Y = f(X) + \epsilon$$

1. Irreducible: ϵ
2. Reducible: $\hat{f}(X)$ is not similar to $f(X)$

Bias Variance Tradeoff

Bias Variance Tradeoff

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon).$$

Bias Variance Tradeoff

$$\boxed{E\left(y_0 - \hat{f}(x_0)\right)^2} = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon).$$

Expected MSE at x_0 if we repeatedly estimated f(x) with different training sets

Bias Variance Tradeoff

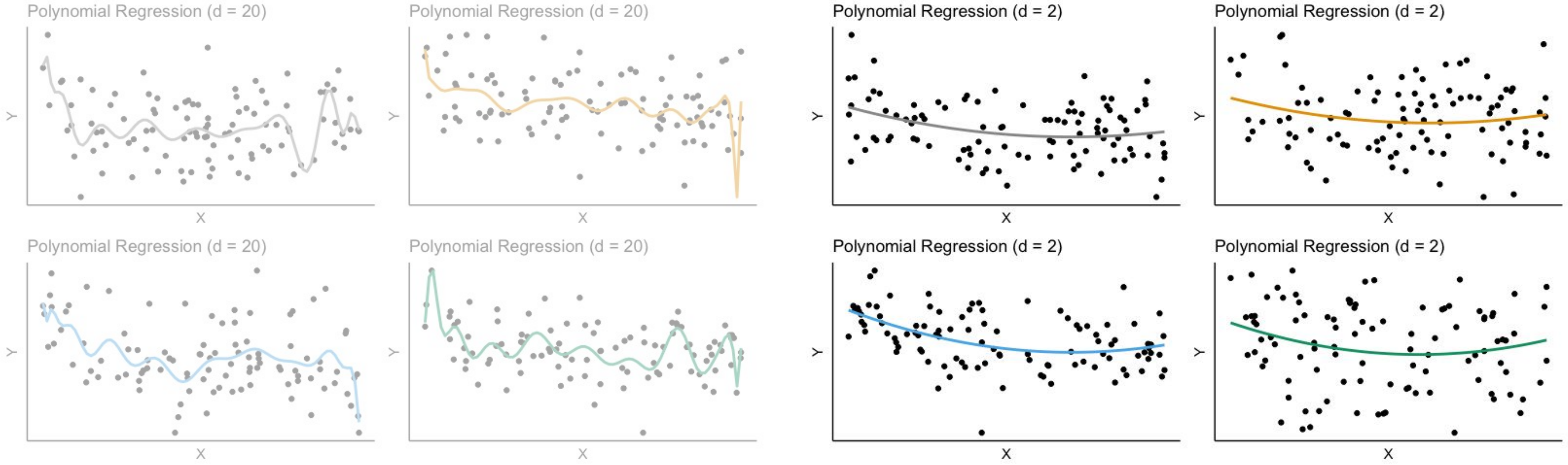
$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \underbrace{\boxed{\mathrm{Var}(\epsilon).}}_{\text{Irreducible Error}}$$

Bias Variance Tradeoff

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \boxed{\mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2} + \mathrm{Var}(\epsilon).$$

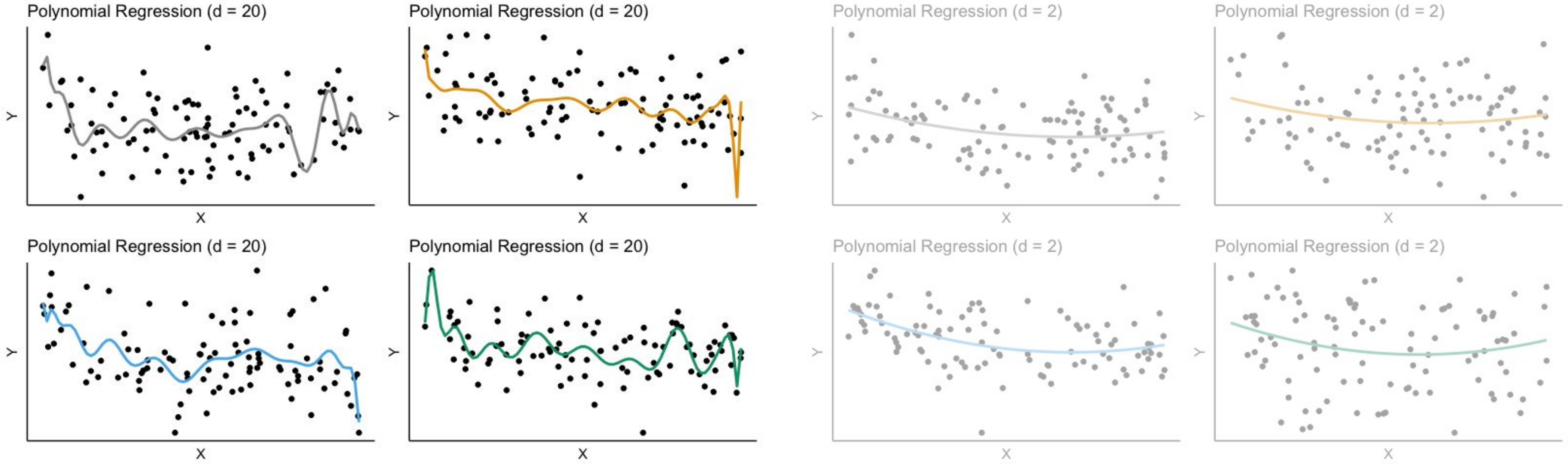
Bias Variance Tradeoff

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \boxed{\mathrm{Var}(\hat{f}(x_0))} + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon).$$



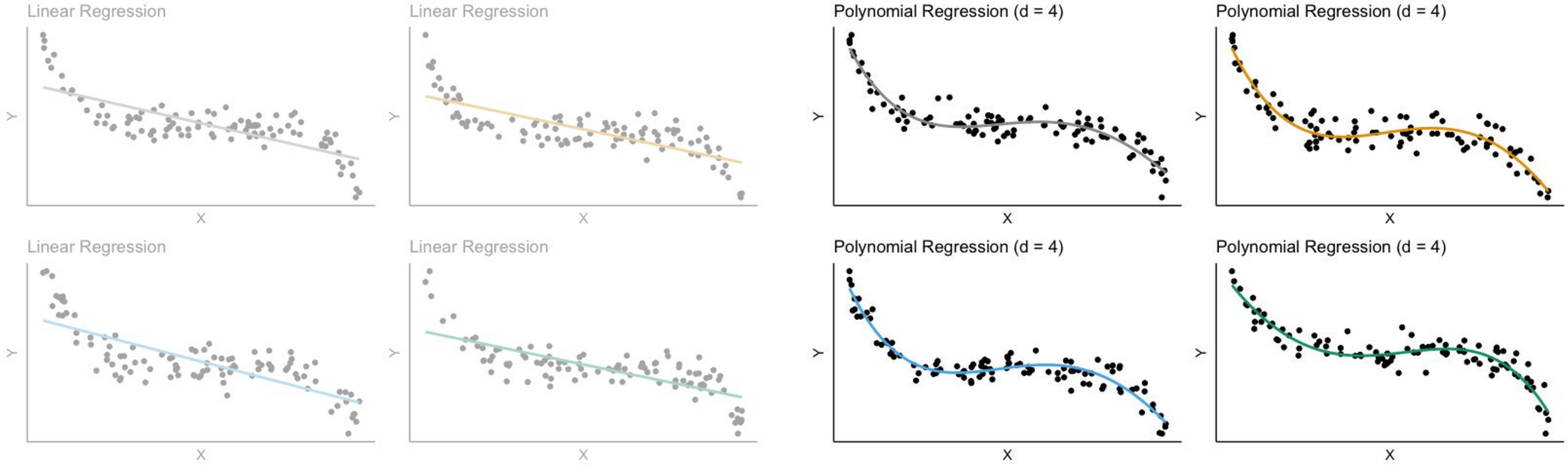
Bias Variance Tradeoff

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \boxed{\mathrm{Var}(\hat{f}(x_0))} + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon).$$



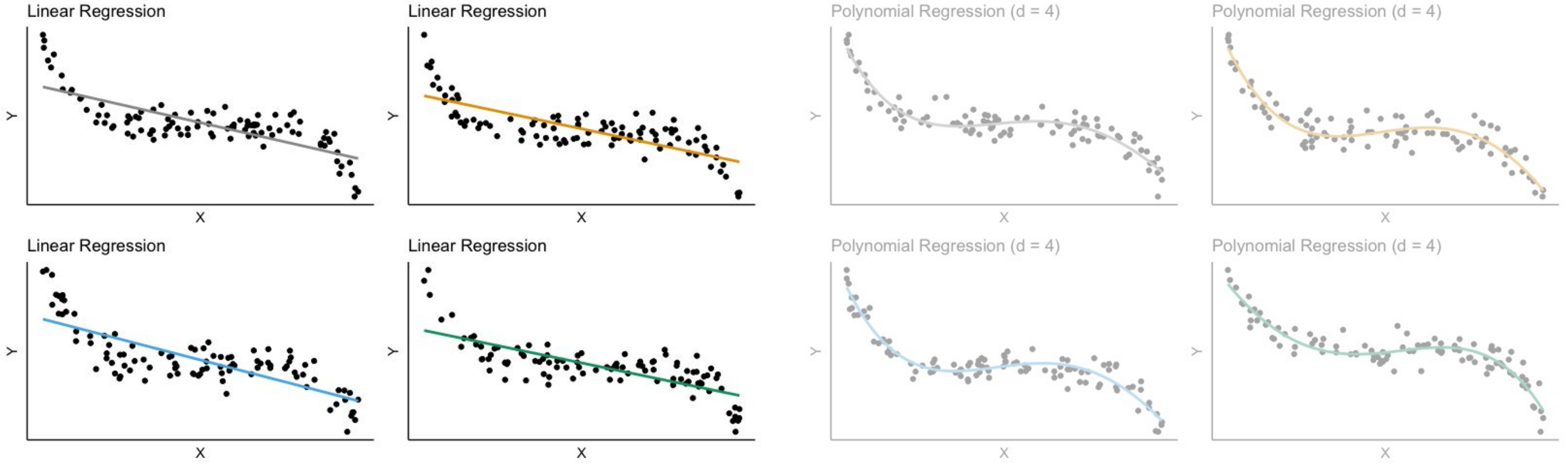
Bias Variance Tradeoff

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \mathrm{Var}(\hat{f}(x_0)) + \boxed{[\mathrm{Bias}(\hat{f}(x_0))]^2} + \mathrm{Var}(\epsilon).$$



Bias Variance Tradeoff

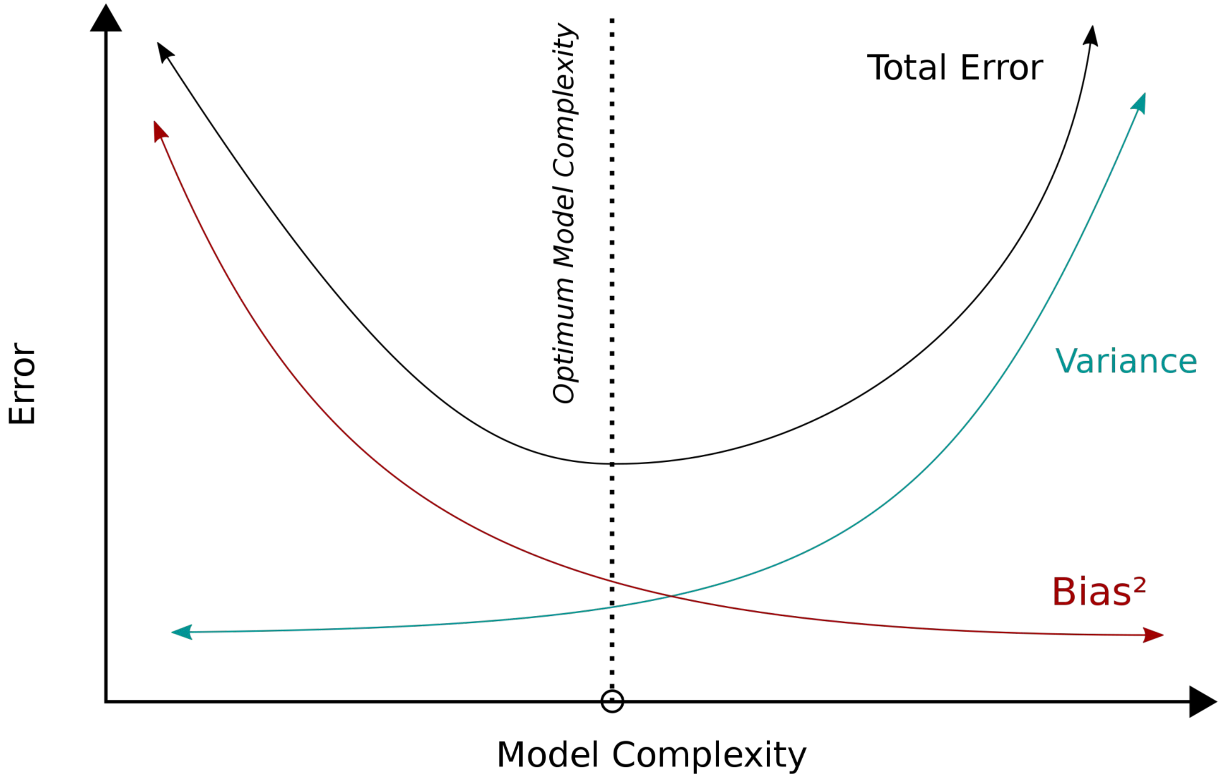
$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \mathrm{Var}(\hat{f}(x_0)) + \boxed{[\mathrm{Bias}(\hat{f}(x_0))]^2} + \mathrm{Var}(\epsilon).$$

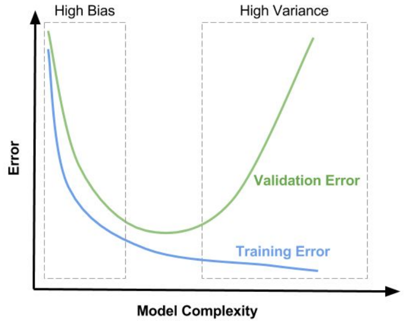


Bias Variance Tradeoff

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \boxed{\mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2} + \mathrm{Var}(\epsilon).$$

Bias Variance Tradeoff



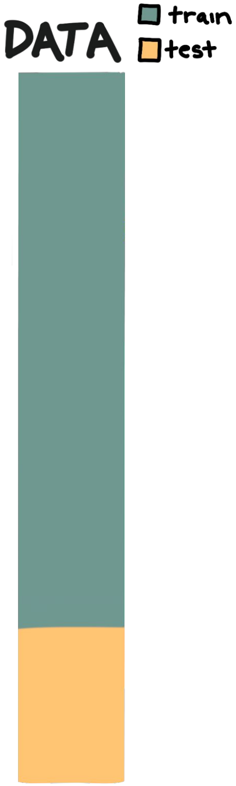


Model Validation

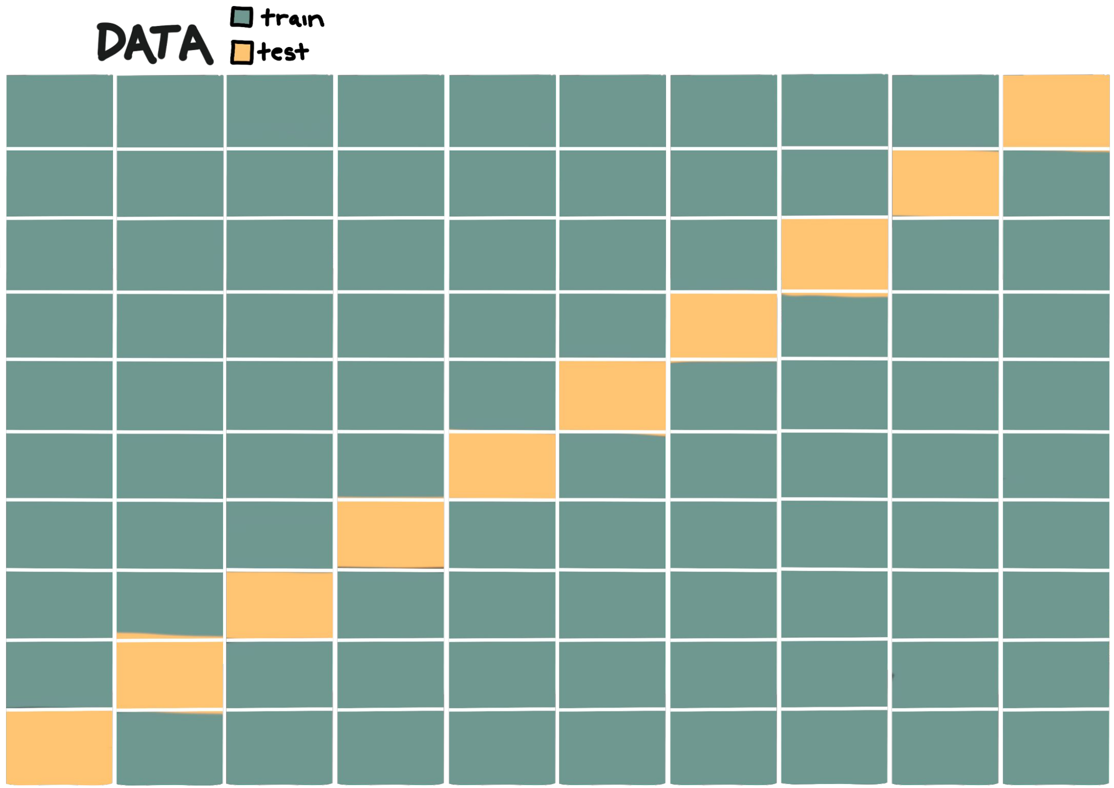
Validation

1. **Split** Data
2. Train Model on **Training Set**
3. Evaluate **Train and Test Set**

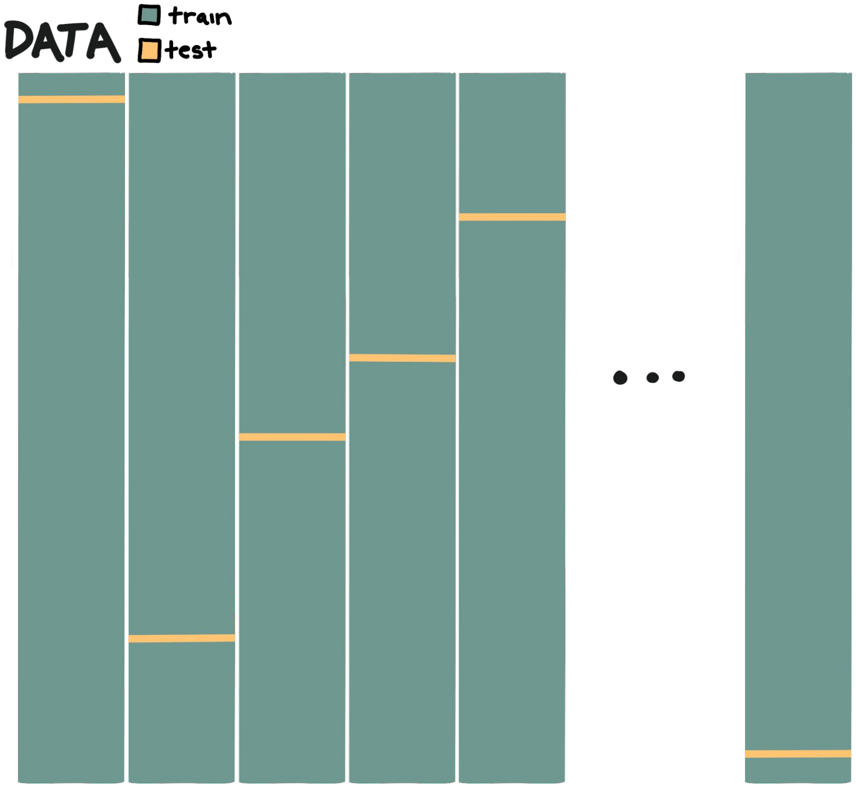
Validation (Train Test Split)



Cross Validation (KFold)



Cross Validation (Leave One Out)

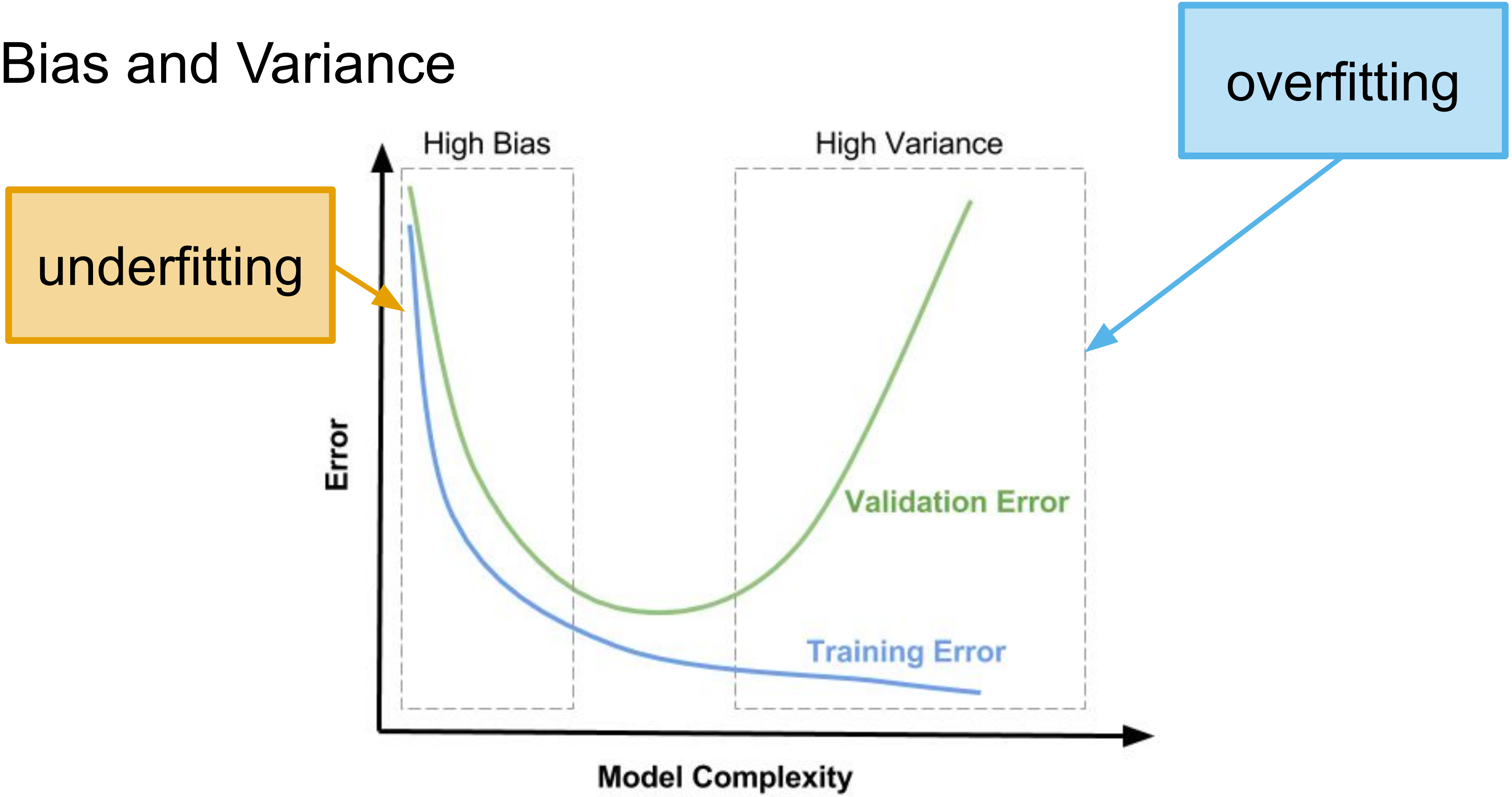


Validation

How to decide

- Size of your Dataset (rows AND columns)
- Computational Expense

Bias and Variance



Regularization

Review: Loss Functions

MINIMIZE:

$$\sum \big(\overbrace{x_i - \hat{x}_i}^{\text{how off we were}}\big)^2$$

true value

model's guess

Ridge

MINIMIZE:

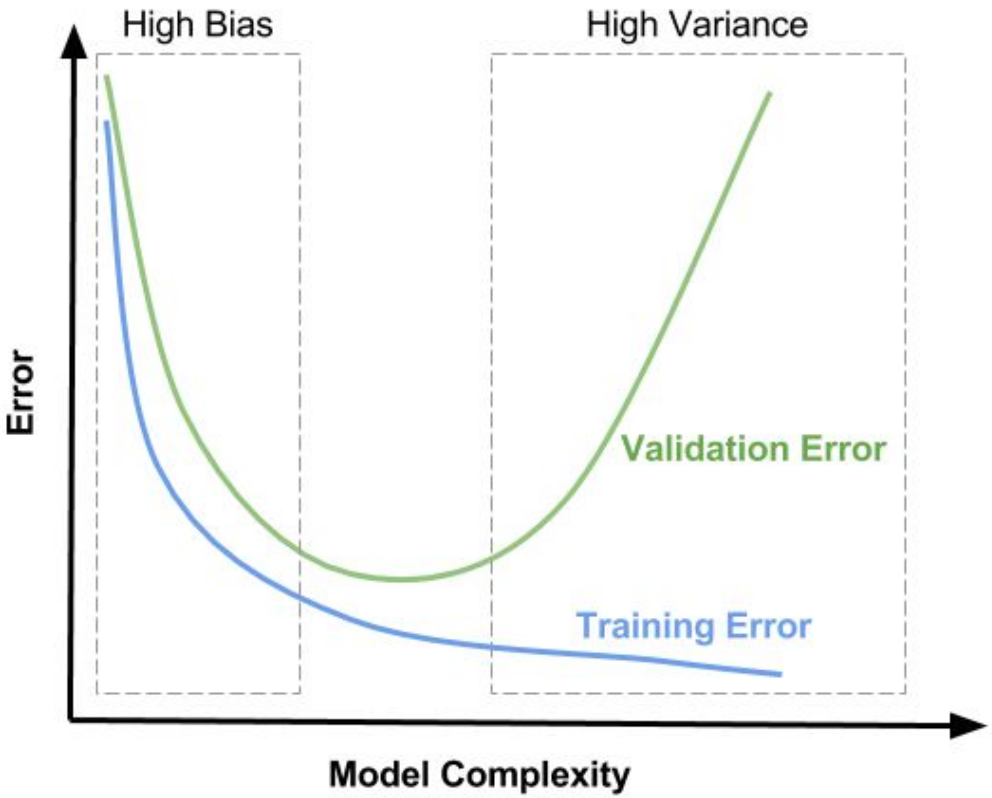
$$\sum \Big(\overbrace{\underbrace{x_i}_{\text{true value}} - \underbrace{\hat{x}_i}_{\text{model's guess}}}^{\text{how off we were}} \Big)^2 + \underbrace{\lambda}_{\text{how HARSHLY we penalize}} \sum \underbrace{B_j^2}_{\text{how big the coefs are}}$$

LASSO

MINIMIZE:

$$\sum \left(\overbrace{\underset{\text{true value}}{x_i} - \underset{\text{model's guess}}{\hat{x}_i}}^{\text{how off we were}} \right)^2 + \underset{\text{how HARSHLY we penalize}}{\lambda} \sum \underset{\text{how big the coefs are}}{|B_j|}$$

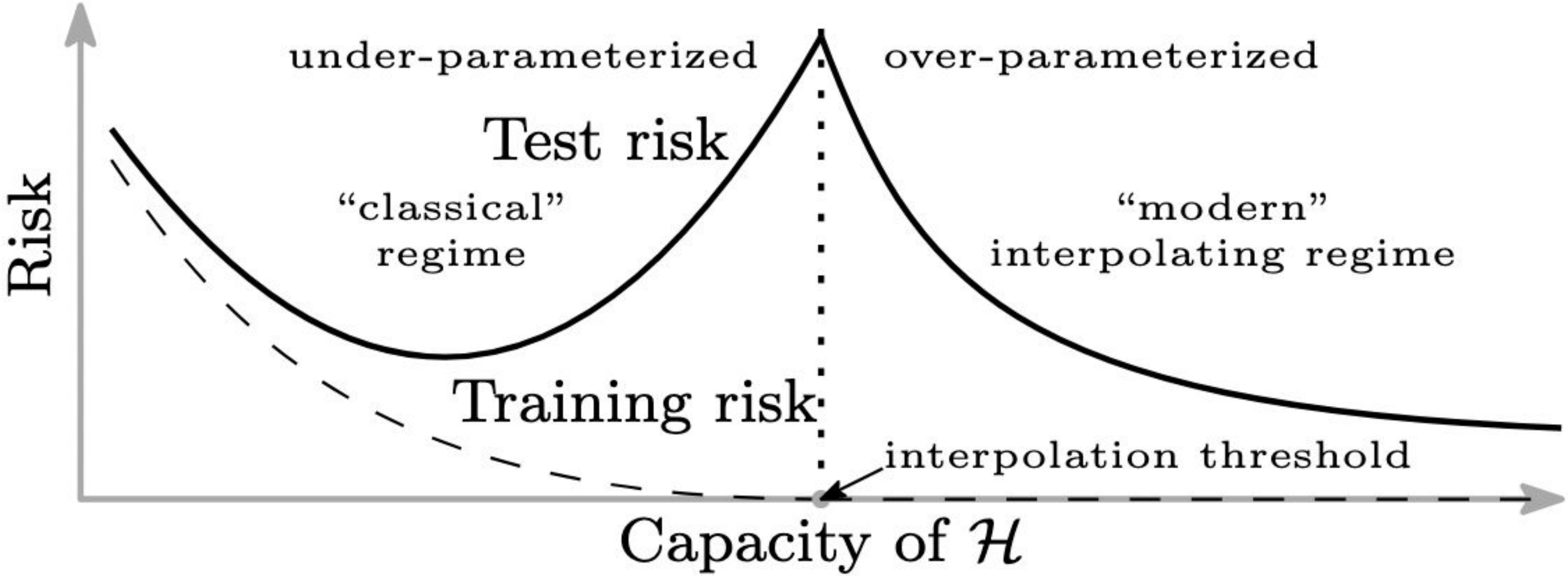
Bias and Variance



Double Descent

Everything I’ve Just
Taught You is
Wrong

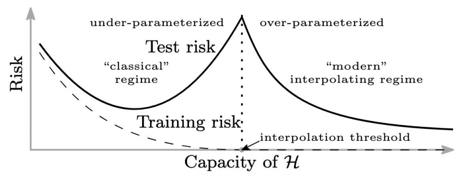
Double Descent



From: https://medium.com/mlearning-ai/double-descent-8f92dfdc442f

Or is it?

From: https://medium.com/mlearning-ai/double-descent-8f92dfdc442f



Double Descent

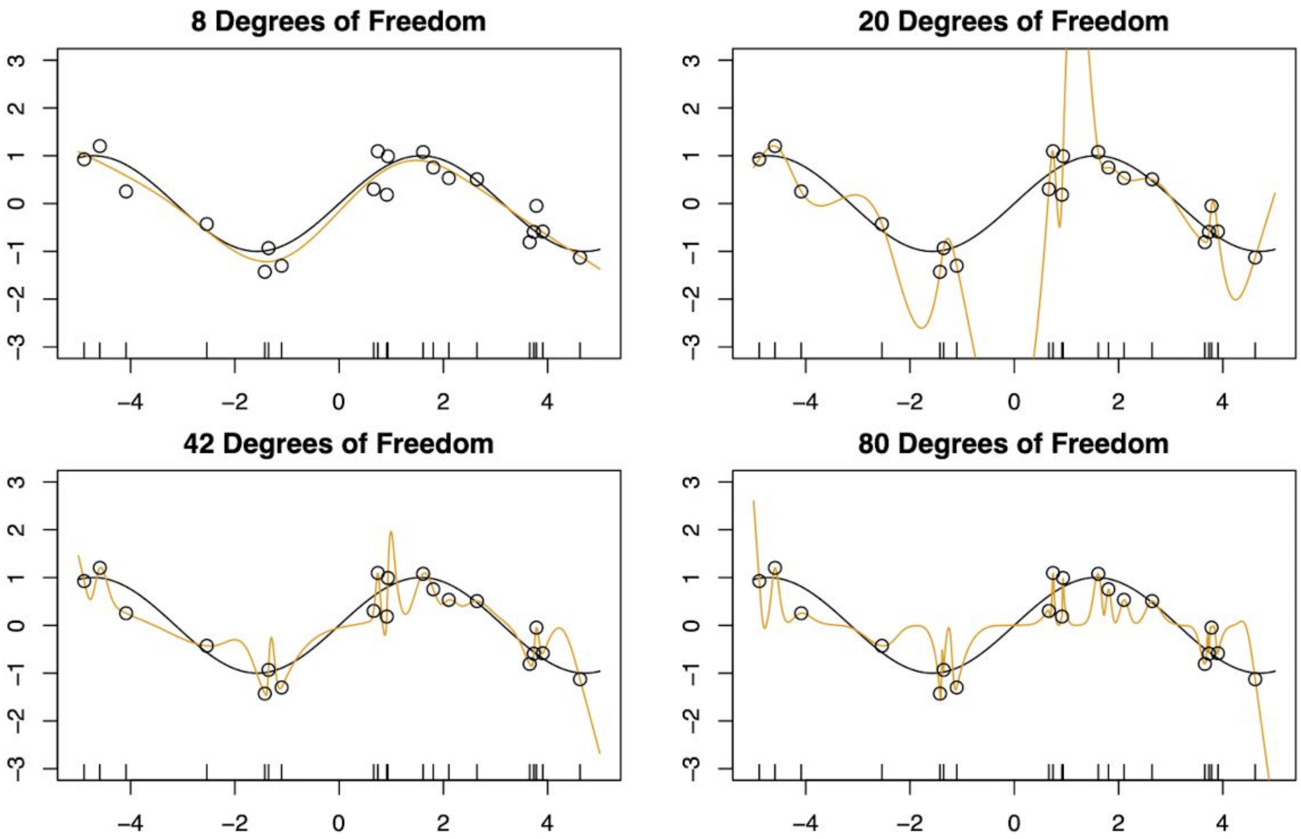


Image from: Introduction to Statistical Learning with applications in Python (Figure 10.21)