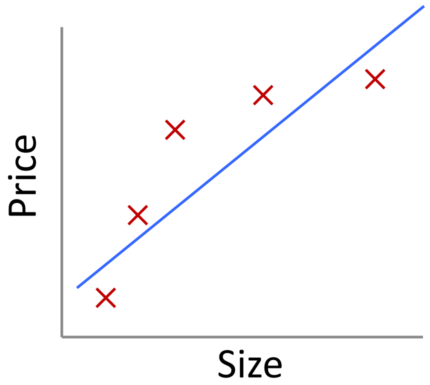
CS M146: Introduction to Machine Learning
Bias-Variance Tradeoff

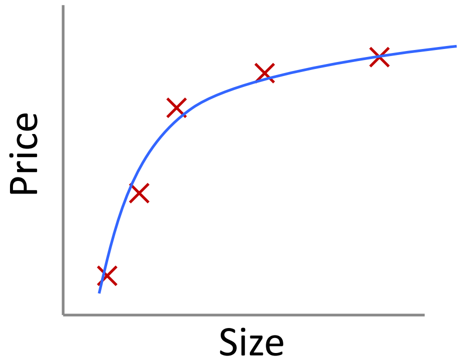
Aditya Grover

https://aditya-grover.github.io/ @adityagrover_

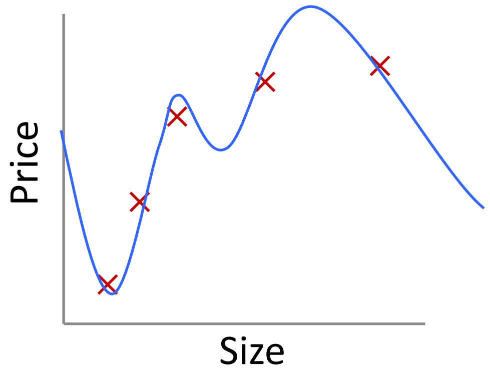
Recap: Underfitting vs Overfitting



$$\theta_0 + \theta_1 x$$



$$\theta_0 + \theta_1 x + \theta_2 x^2$$

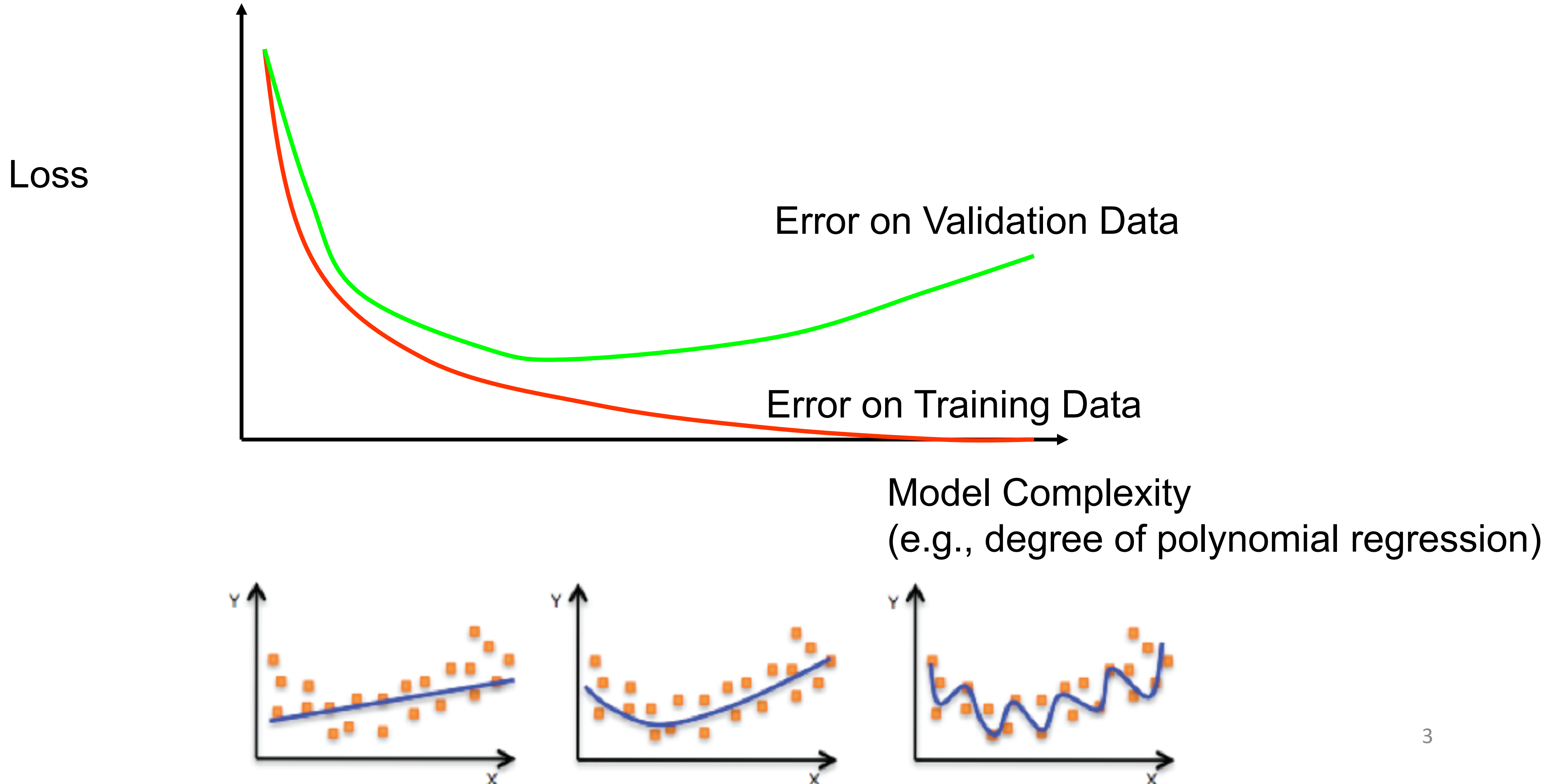


$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

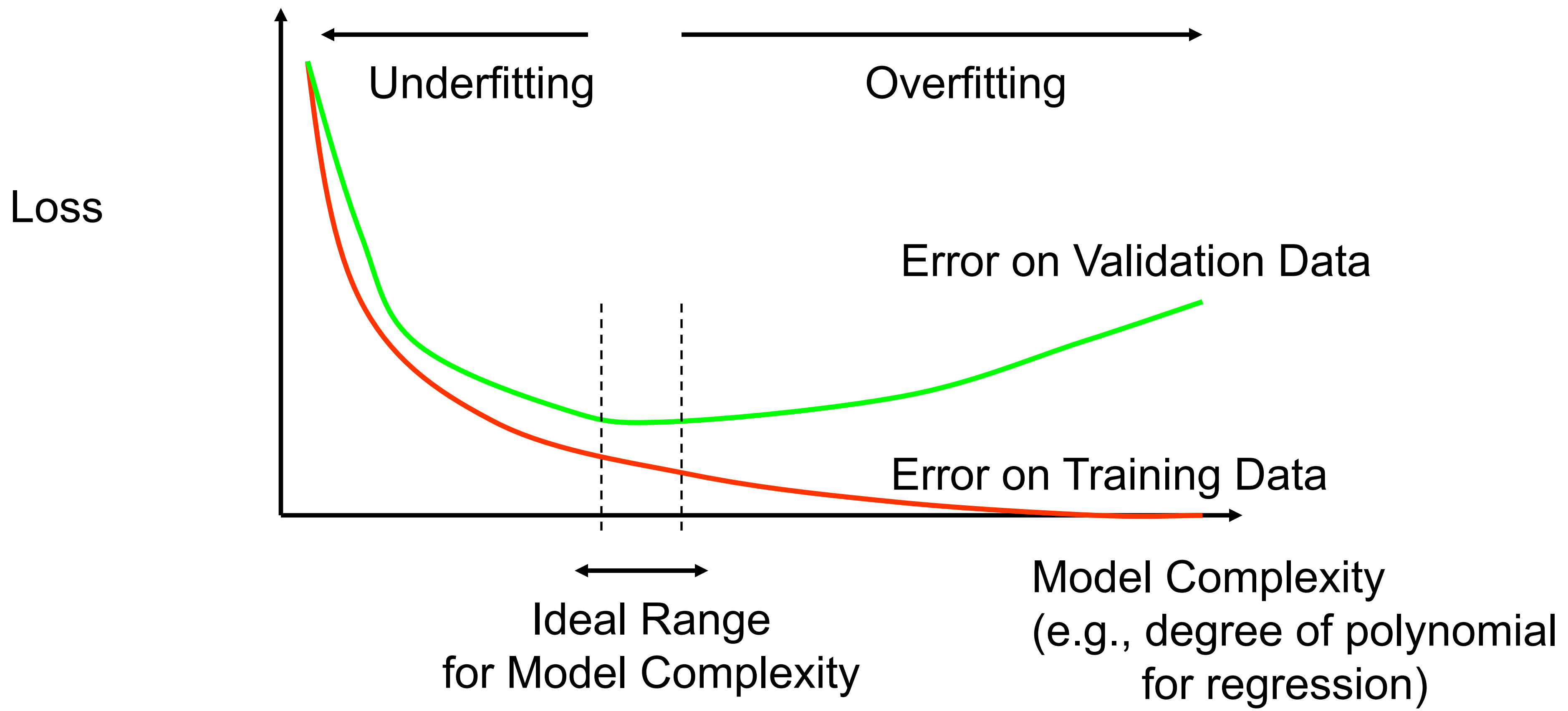
underfitting

overfitting

Recap: Model Complexity Curves

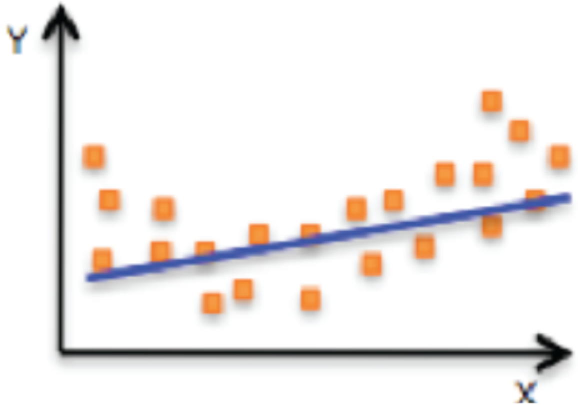


Recap: Model Complexity Curves

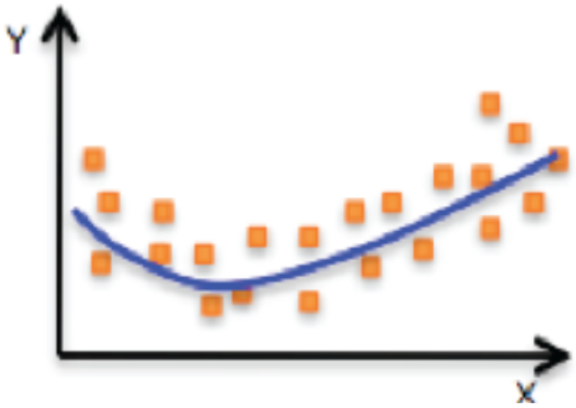


Underfitting and overfitting show very different behaviors on training and validation data

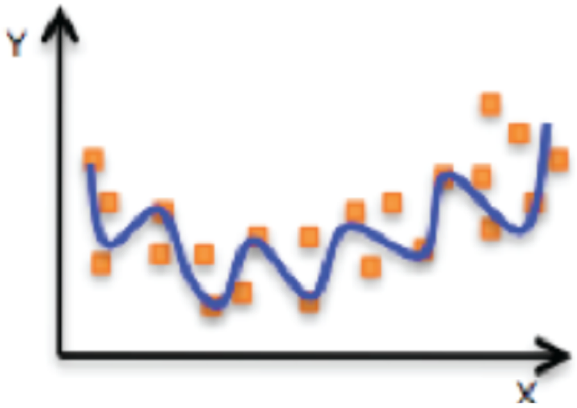
Recap: Detecting Underfitting and Overfitting



Case 1



Case 2



Case 3

	Case 1	Case 2	Case 3
Training loss:	High	Medium	Low
Validation loss:	High	Medium	High
	Underfitting	**Good fit**	**Overfitting**

Can we provide statistical definitions?

Bias-Variance Tradeoff

- Bias of an estimator: Difference between an estimator's expected value and true value
  - ML: How far are the model's predictions (in expectation) from the true predictions?
  - High bias → underfitting

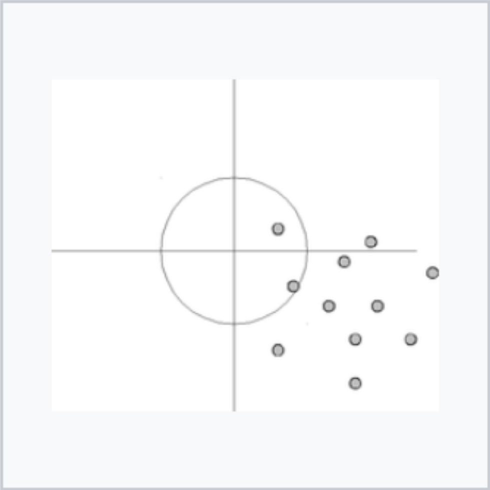
- Variance of an estimator:
  - ML: If we train a ML model on different training sets of size n, how much do the predictions vary on a test set?
  - High variance → overfitting

Visual Analogy with Throwing Darts

bias low, variance low

bias high,
variance low:

bias low,
variance high:



bias high,
variance high:

Error Decomposition

- (Stated without proof)

For regression, the expected squared error for any hypothesis can be decomposed into three components

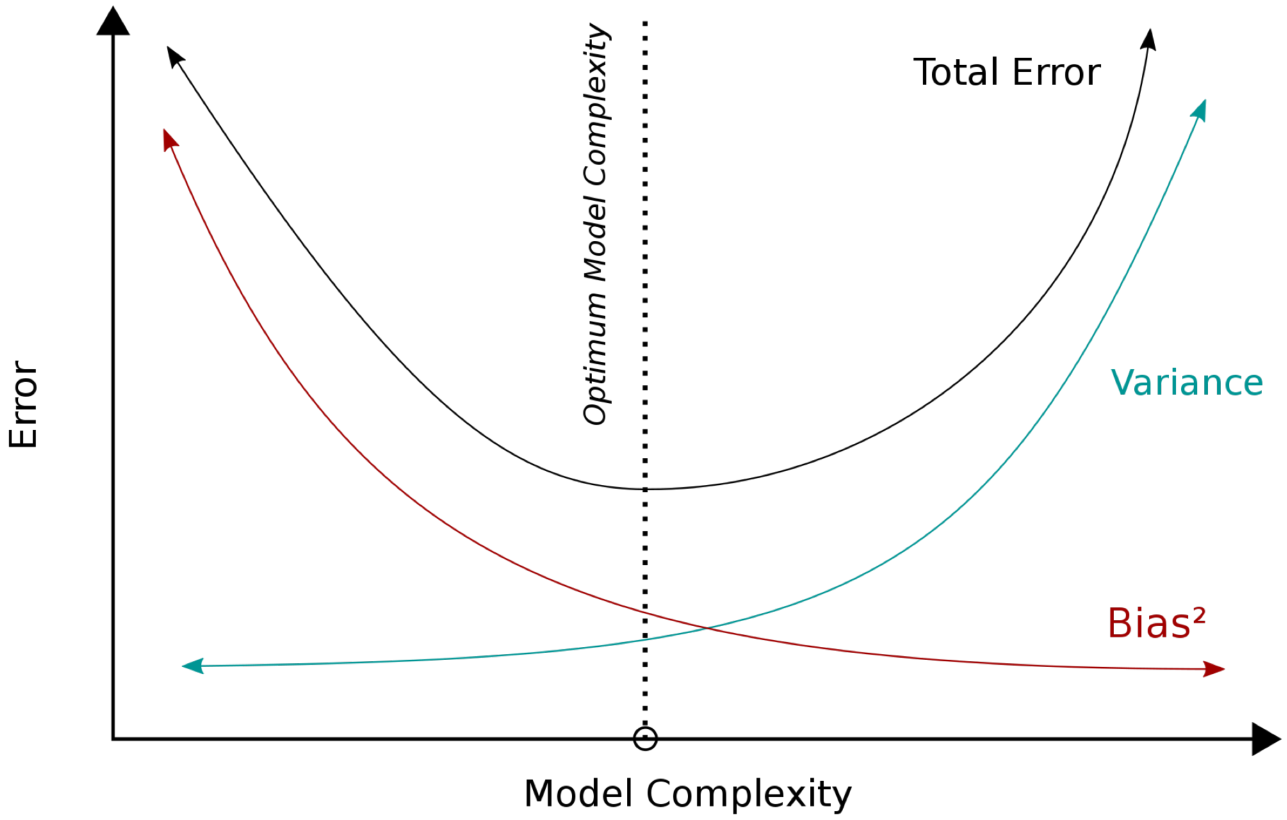
Expected Squared Error:

Noise in the training data

+ Bias^2

+ Variance

Error Decomposition



Fixing Errors

Formally, we have:

- noisy data → irreducible error

  Fix: find another source of high-quality data

- underfitting → high bias

  Fix: select a more complex hypothesis class

- overfitting → high variance

  Fix: increase size of dataset

  or reduce complexity of hypothesis class

Learning Curves

- **Model complexity curves** plot performance (y-axis) as a function of the **complexity of different models** (x-axis)

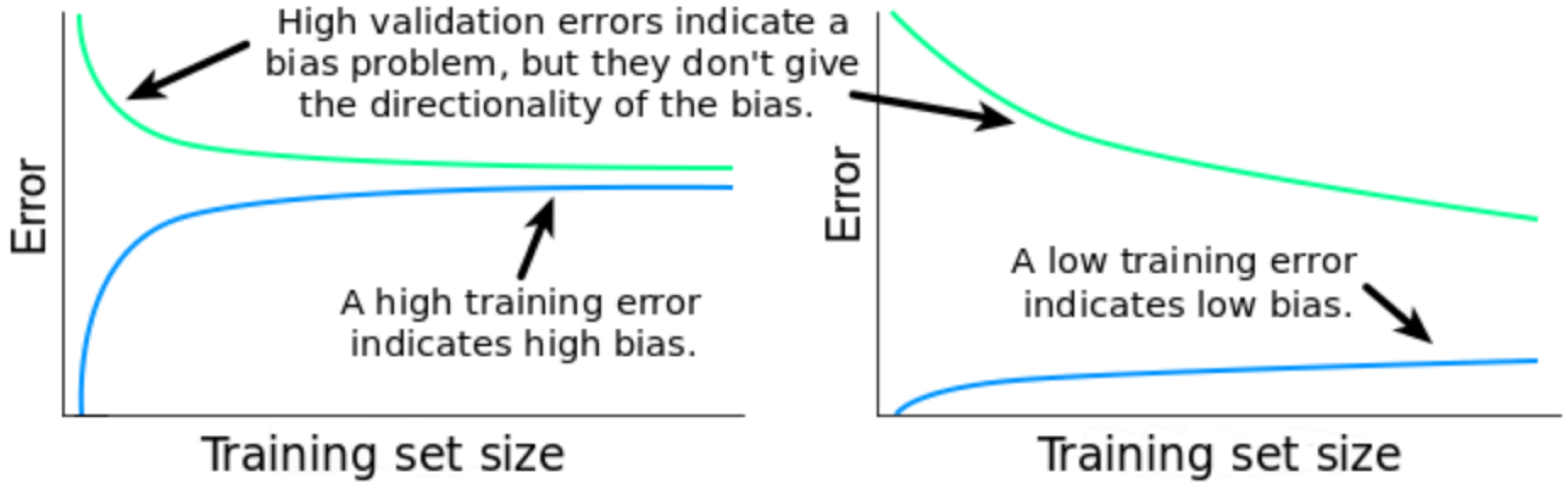
- For a given model, a **learning curve** plots performance (y-axis) as a function of the **size of the training data** (x-axis)

- **Intuition:**

- Magnitude of training error indicates bias
- Gap in training and validation error indicates variance

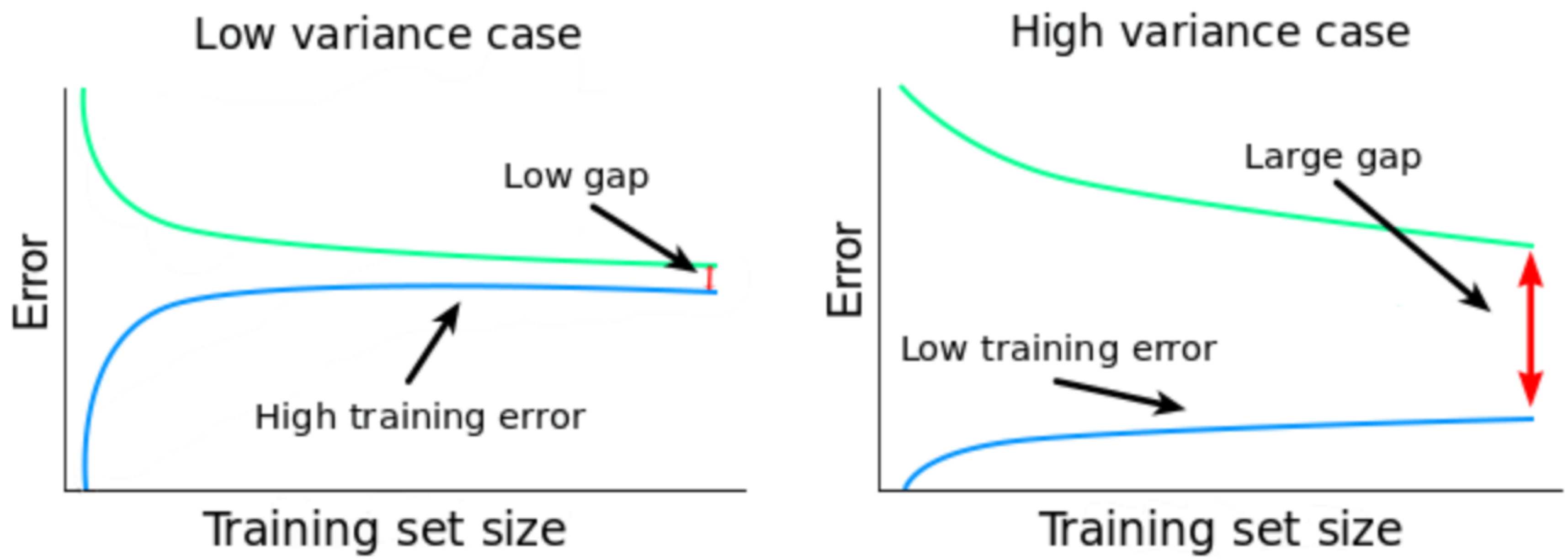
Learning Curves: Assessing Bias

High bias | Low bias



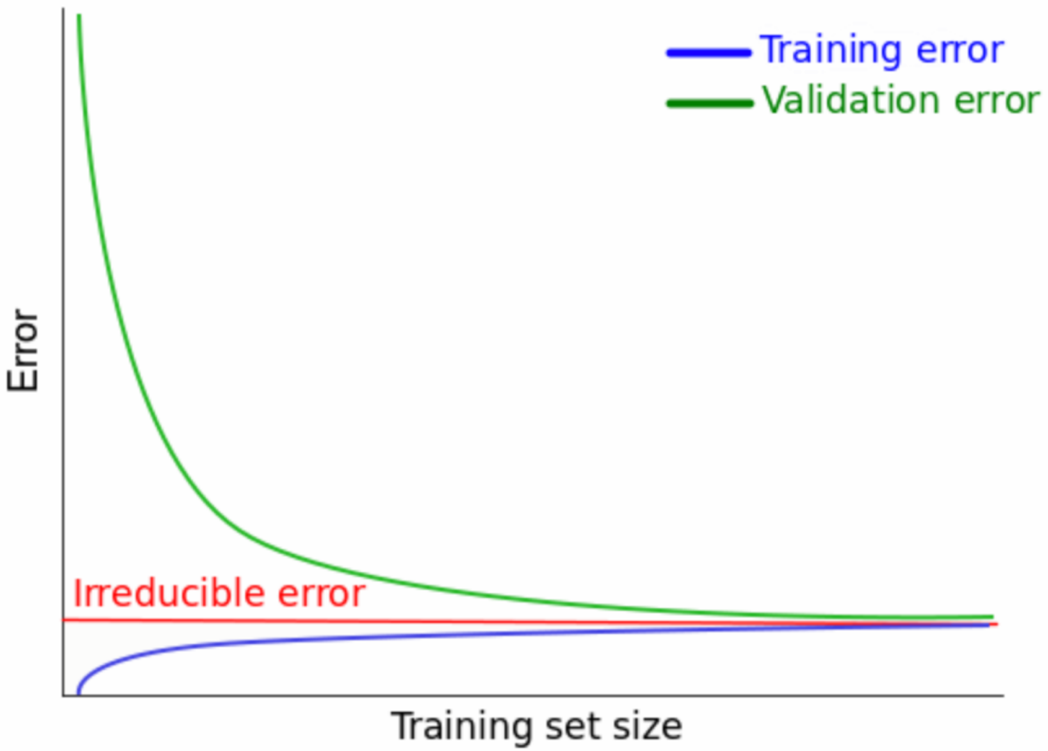
Assumption: Irreducible error is not very high

Learning Curves: Assessing Variance



Learning Curves

- Idealized learning curve



Summary

Bias-Variance Tradeoff

- Another formal perspective on overfitting and underfitting
- Can be detected in practice via learning curves