Bias-Variance Trade-off

Chandranath Adak

Dept. of CSE, IIT Patna

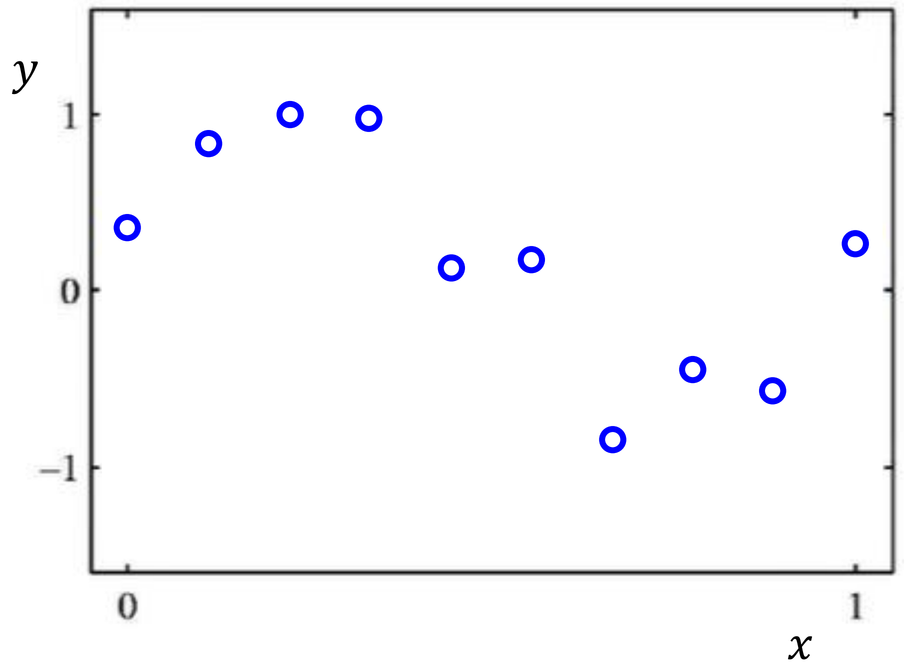
Outline

- Bias
- Variance

Linear Regression: Polynomial Curve Fitting

Objective:
given data set consists of blue dots,
fit into a polynomial of $\sum_{j=0}^{m} w_j\, x^j$ form

$$\hat{y}(x, W) = w_0 + w_1 x + w_2 x^2 + \cdots w_m x^m = \sum_{j=0}^{m} w_j\, x^j$$



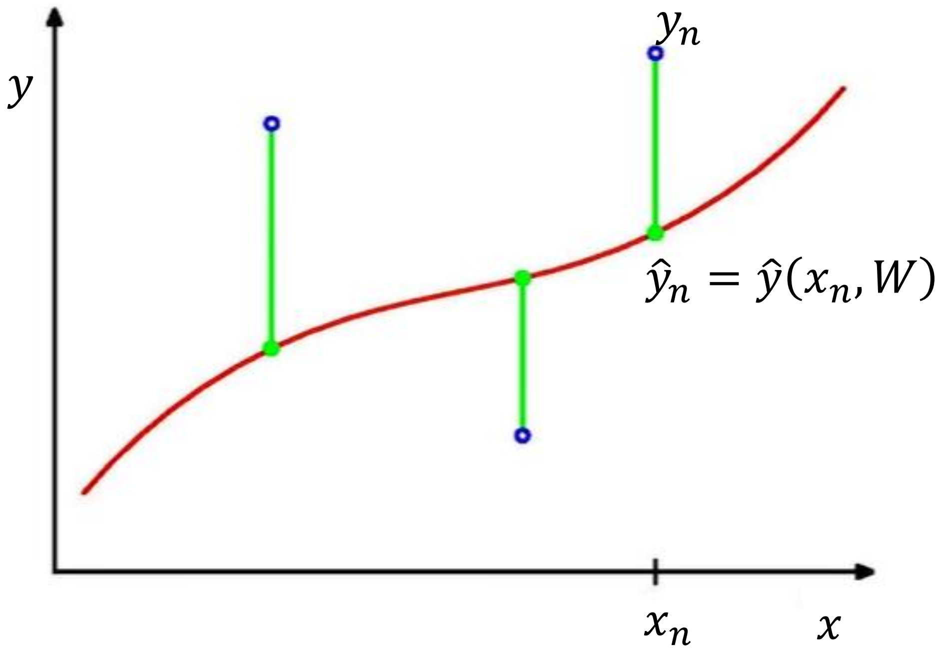
Working principle:
vary the degree of polynomial to see how the fits look like

Fitting Error

What should be the curve?
where error e_i is minimum

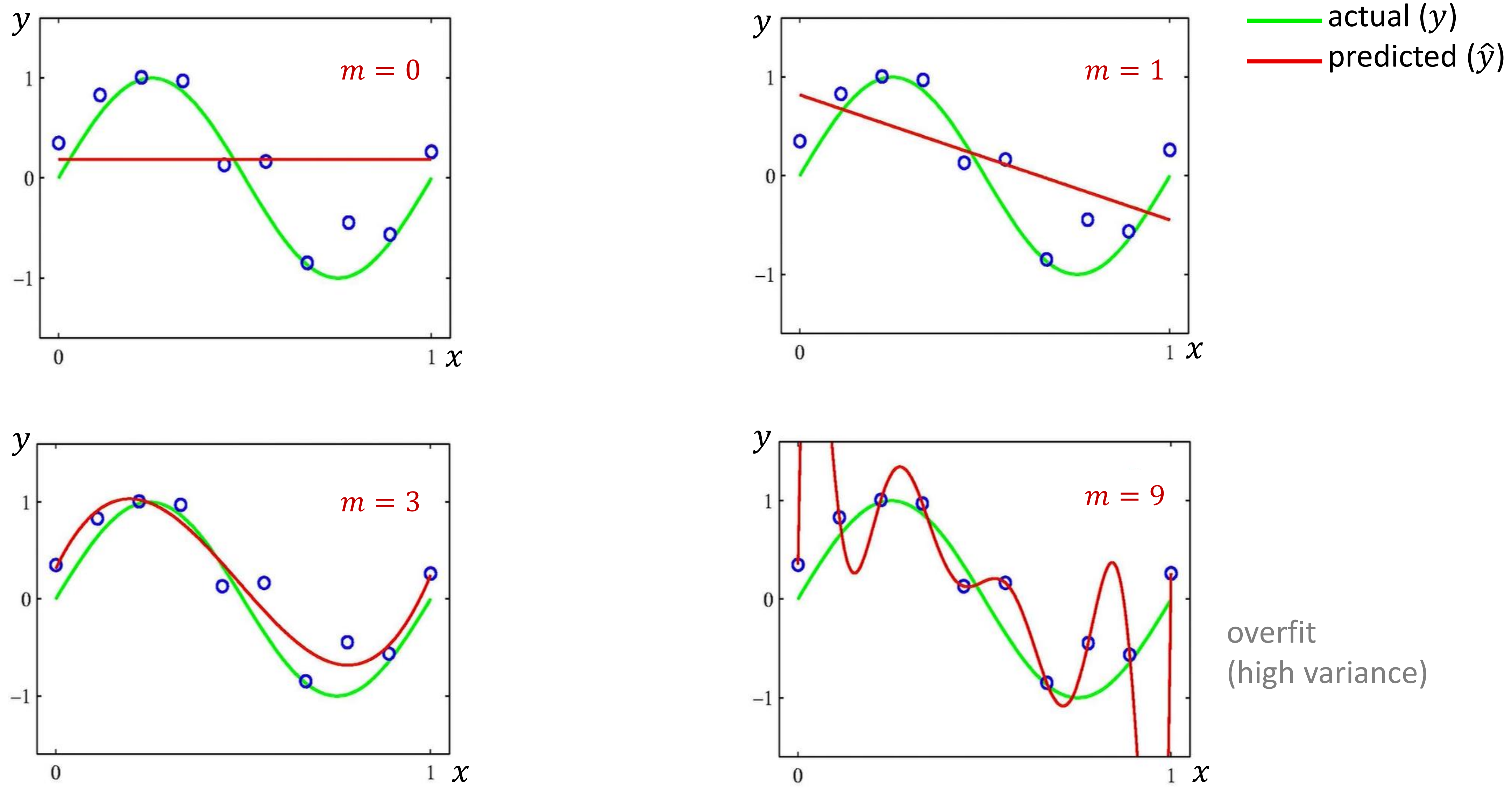
Solution: Least Square method
Objective: To learn W to minimize SS_{Res}

$$SS_{Res} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$



Fitting the Curve with m^{th} order Polynomial

underfit
(high bias)

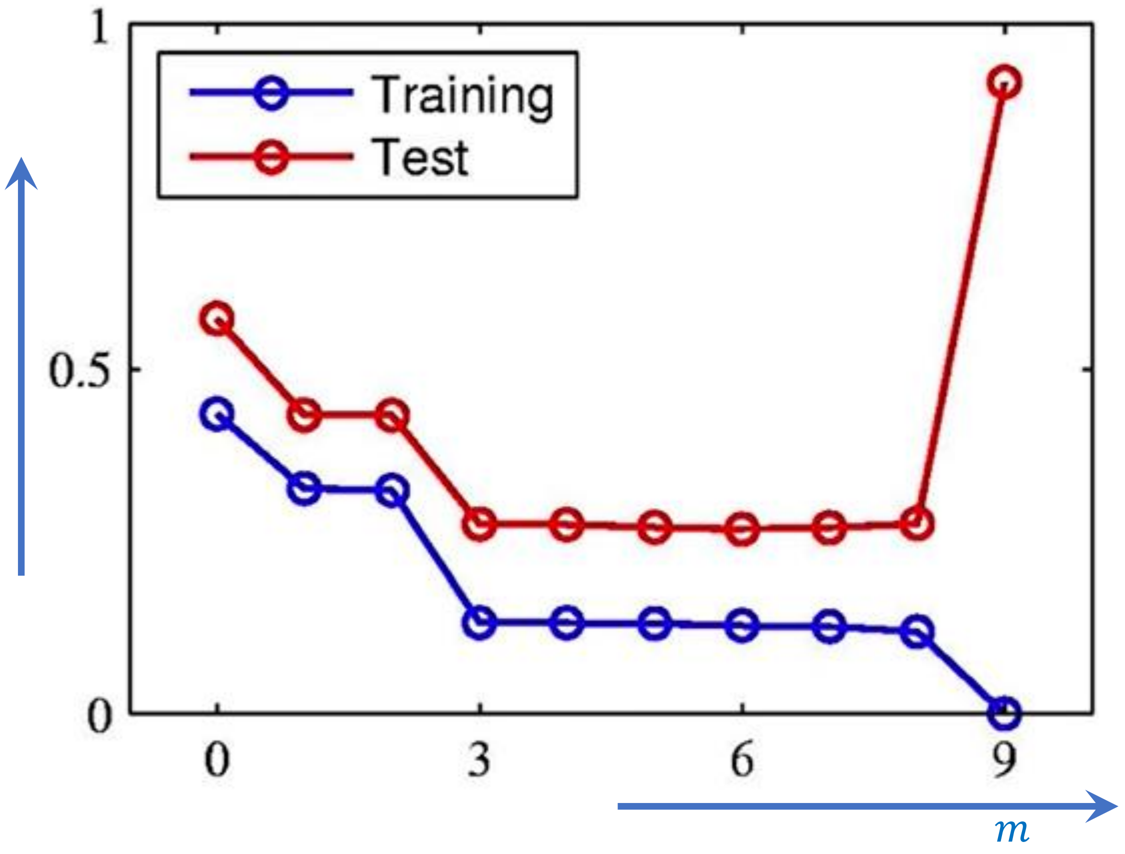


overfit
(high variance)

Training – Testing Curve: Overfitting Intuition

$$E_{RMS} = \sqrt{MSE}$$

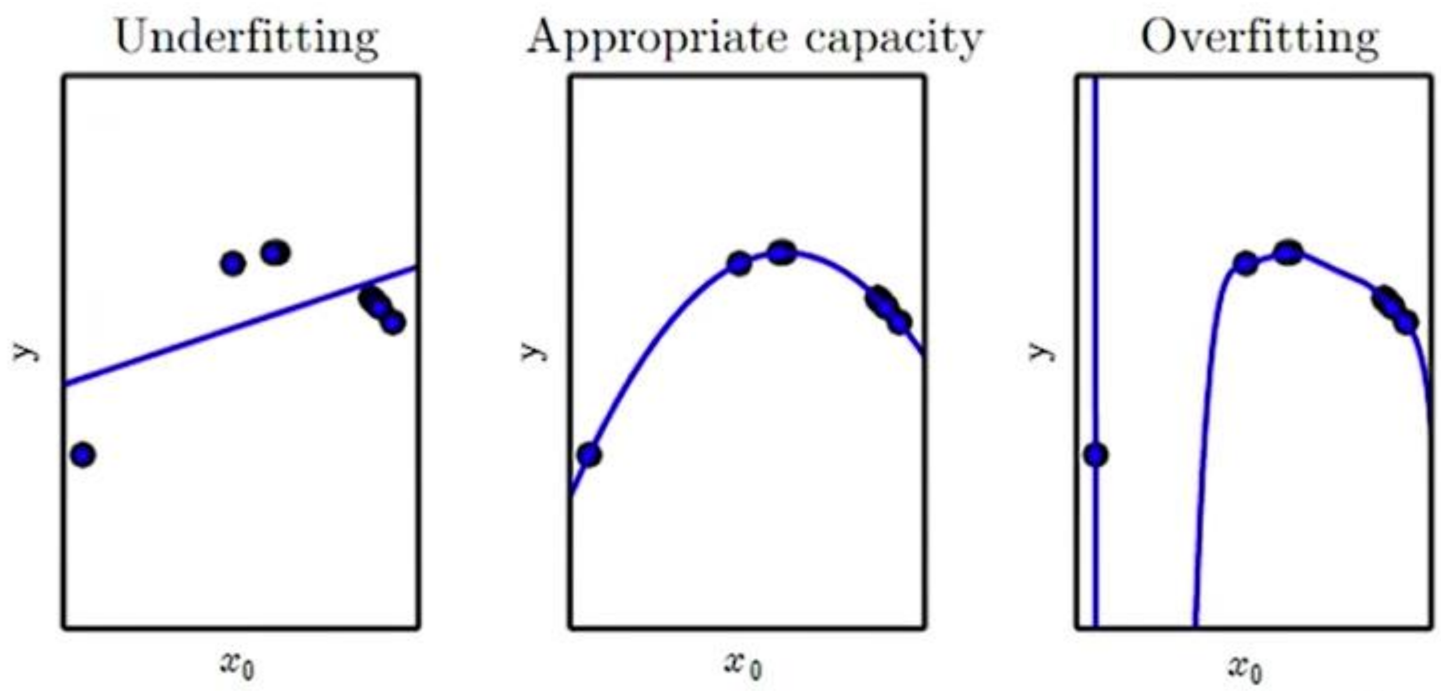
$$= \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$



Capacity: Underfitting vs. Overfitting

High bias

High variance



$$\hat{y} = b + wx$$

$$\hat{y} = b + w_1 x + w_2 x^2$$

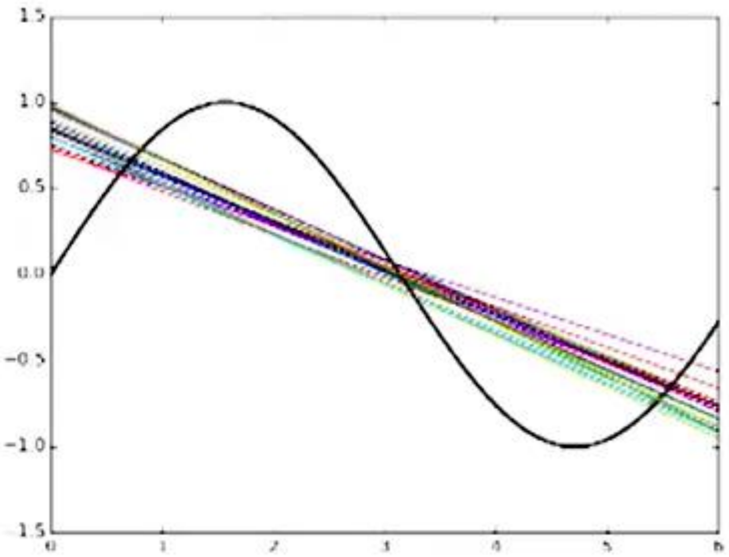
$$\hat{y} = b + \sum_{i=1}^{9} w_i x^i$$

Simple

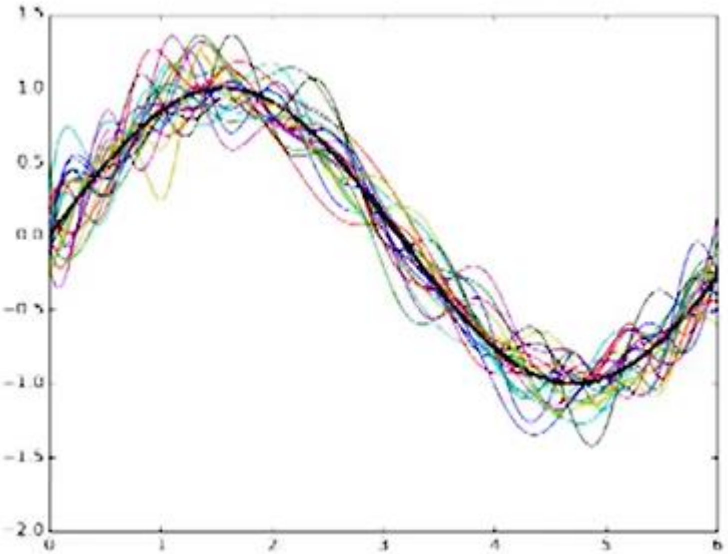
Complex

Underfitting vs. Overfitting

- Training data contains 100 points
- We sample 25 points from training data and train a simple and complex model.
- We repeat the process 'k' times to train each model.
- Simple models trained on different samples of the don't differ much from each other.
  However, they vary far from the true sinusoidal curve.
  (underfitting/ high bias)
- Complex models trained on different samples of the data are very different from each other.
  (overfitting/ high variance)



Simple model



Complex model

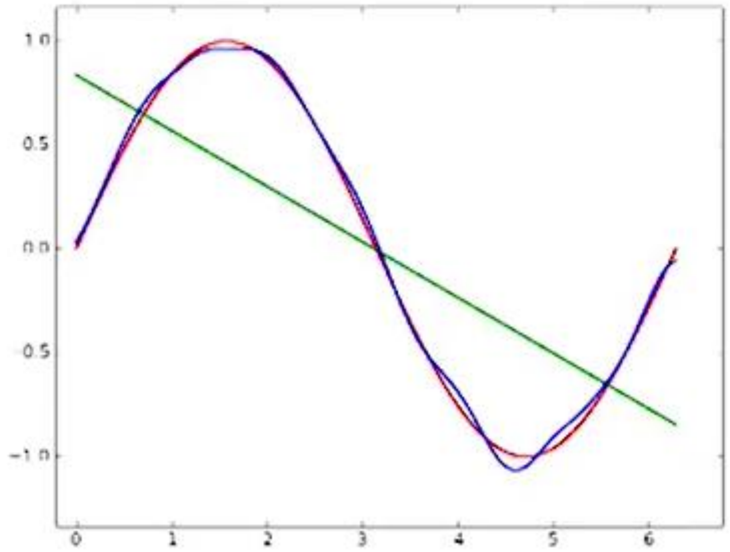
https://www.cse.iitm.ac.in/course_details.php?arg=MTE1

Bias

- $f(x)$ be the true model (sinusoidal here)
- $\hat{f}(x)$ is the estimate of the model (simple or complex)

$$Bias\left(\hat{f}(x)\right) = \mathbb{E}[\hat{f}(x)] - f(x)$$

- $\mathbb{E}[\hat{f}(x)]$ is the expected (or, average) value of the model
- For the simple model the average value (green line) is very far from the true value $f(x)$
- Mathematically, this means the simple model has high bias
- On the other hand, complex model has low bias



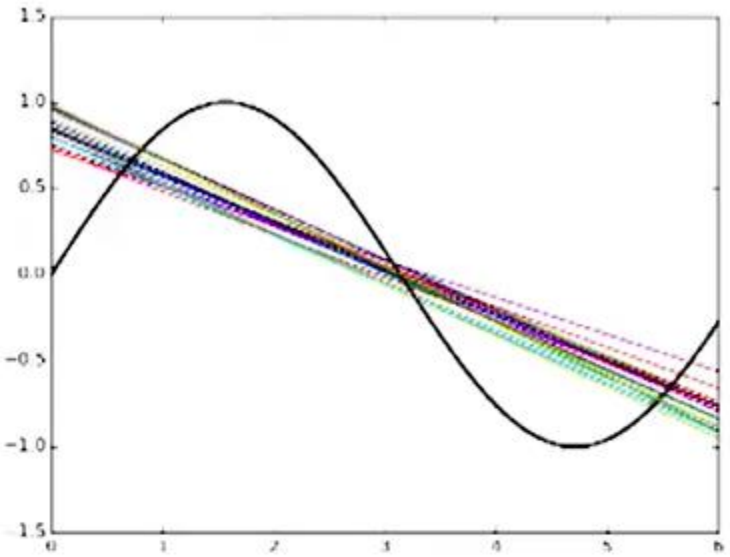
Green : average value of $\hat{f}(x)$ for simple model
Blue : average value of $\hat{f}(x)$ for complex model
Red : true model $f(x)$

Variance

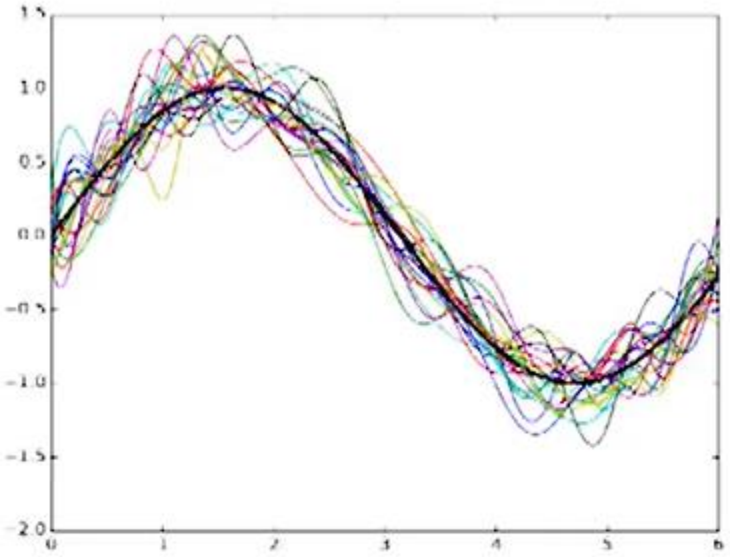
$$Variance\left(\hat{f}(x)\right) = \mathbb{E}\left[\left(\hat{f}(x) - \mathbb{E}[\hat{f}(x)]\right)^2\right]$$

- It tells us how much the different $\hat{f}(x)$'s (trained on different samples of data) differs from each other.

- It can be observed that the
    - simple model has low variance whereas
    - the complex model has high variance

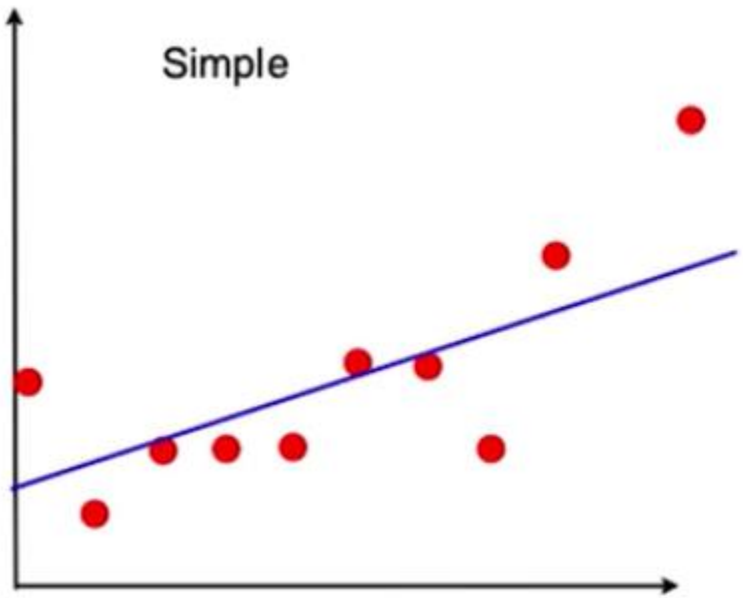


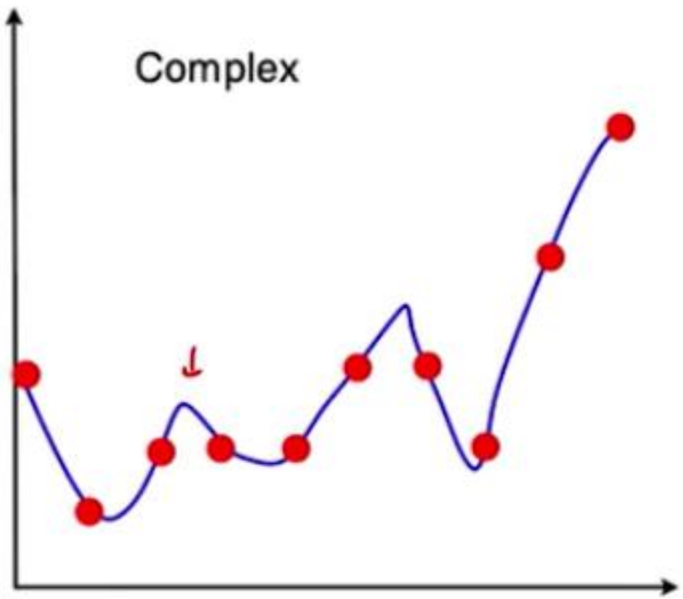
Simple model



Complex model

Bias vs. Variance

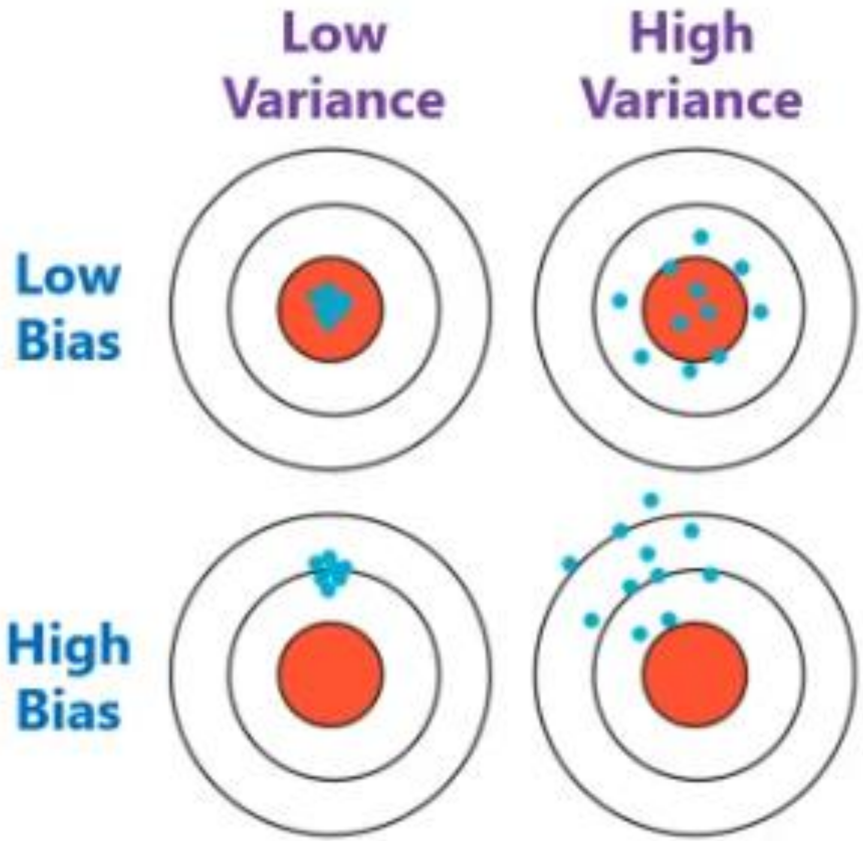


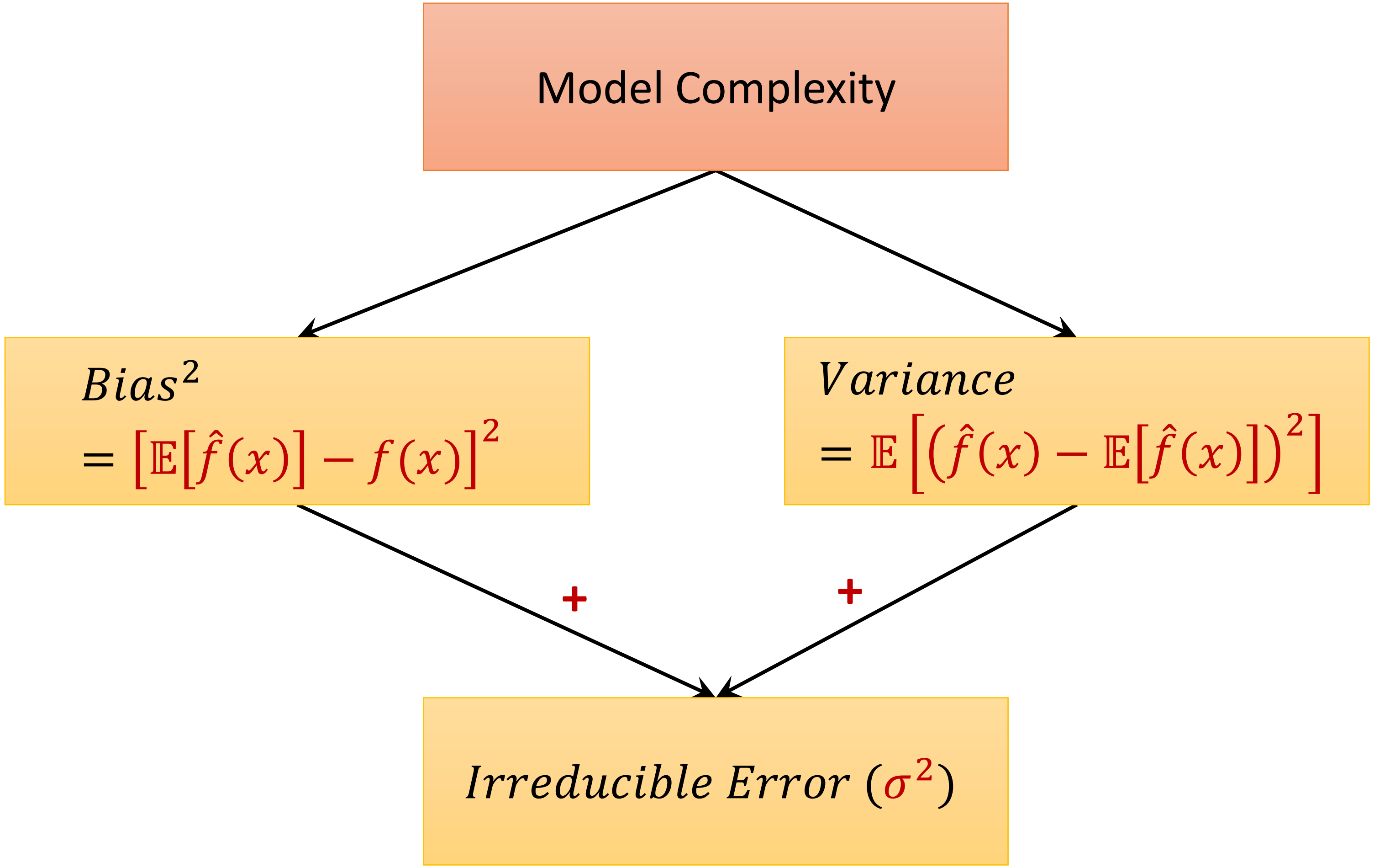


Bias
Underfitting
Insufficient Features
Simple models might have high bias
Complex models might have low bias

Variance
Overfitting
Too many features
Simple models might have low variance
Complex models might have high variance

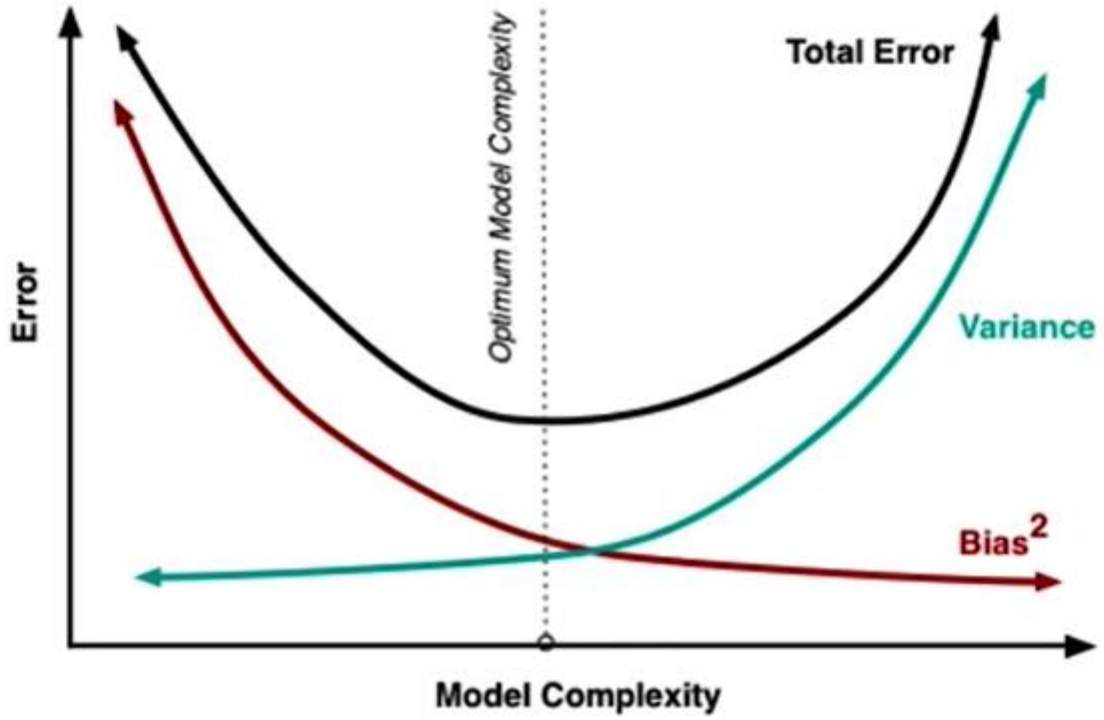
Dart Board Example: Bias vs. Variance

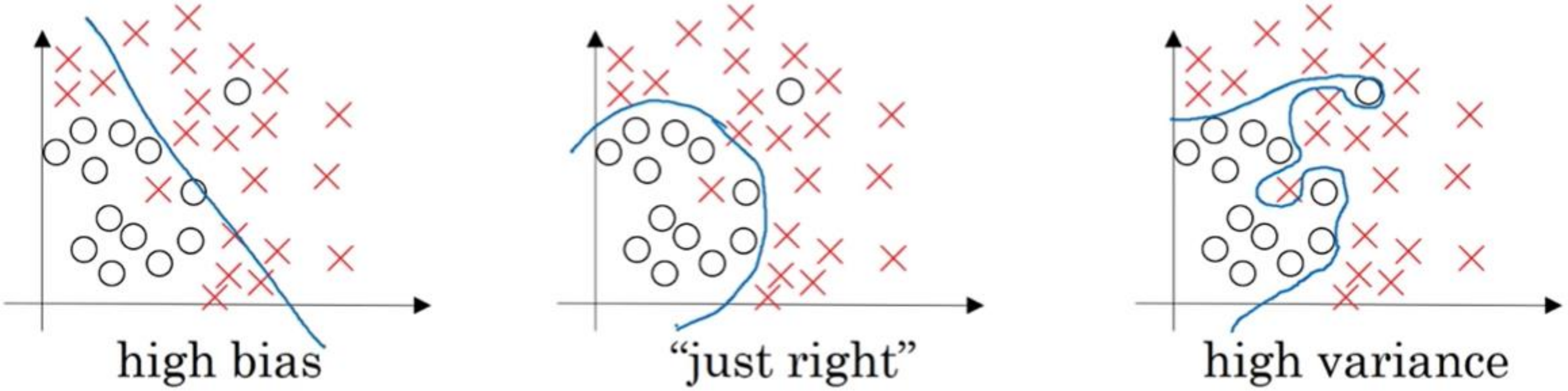




Model Complexity
$Bias^2$
$= [\mathbb{E}[\hat{f}(x)] - f(x)]^2$
$Variance$
$= \mathbb{E}\left[\left(\hat{f}(x) - \mathbb{E}[\hat{f}(x)]\right)^2\right]$
+
+
$Irreducible\ Error\ (\sigma^2)$

Bias-Variance Tradeoff





Underfitting

Overfitting

Training error	1%	15%	15%	0.5%
Testing error	11%	16%	30%	1%
Infer	High variance	High bias	High bias & High variance	Low bias & low variance

Assumption: Human/ optimal (Bayes’) error ≈ 0%

Thank You!