DS 3010

Spring 2026

Wenting Xu

IOWA STATE UNIVERSITY

Module 2: Statistical Decision Tree

Part 1: Bias Variance Tradeoff

Wenting Xu

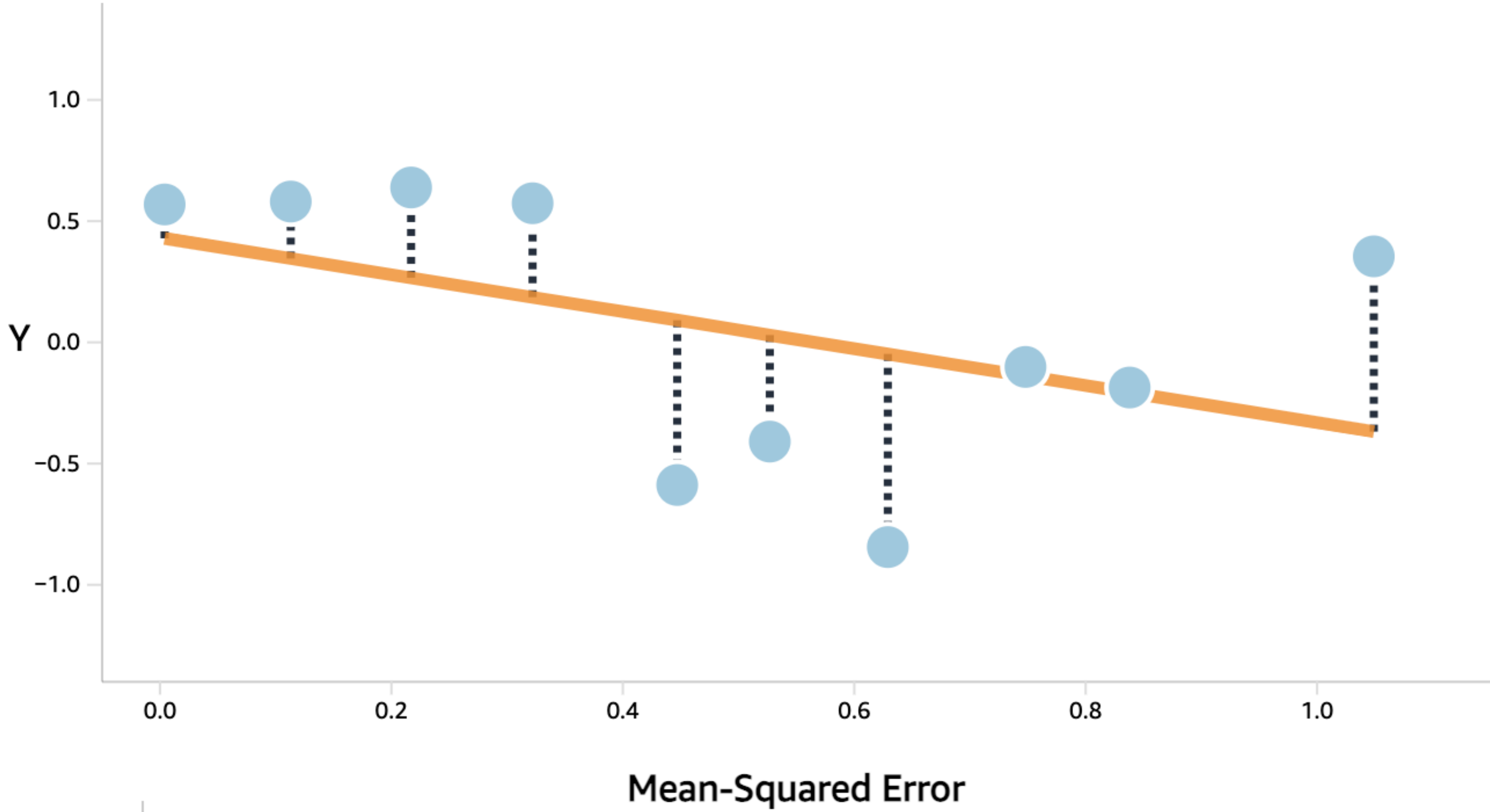
Iowa State University

Copyright and Attribution

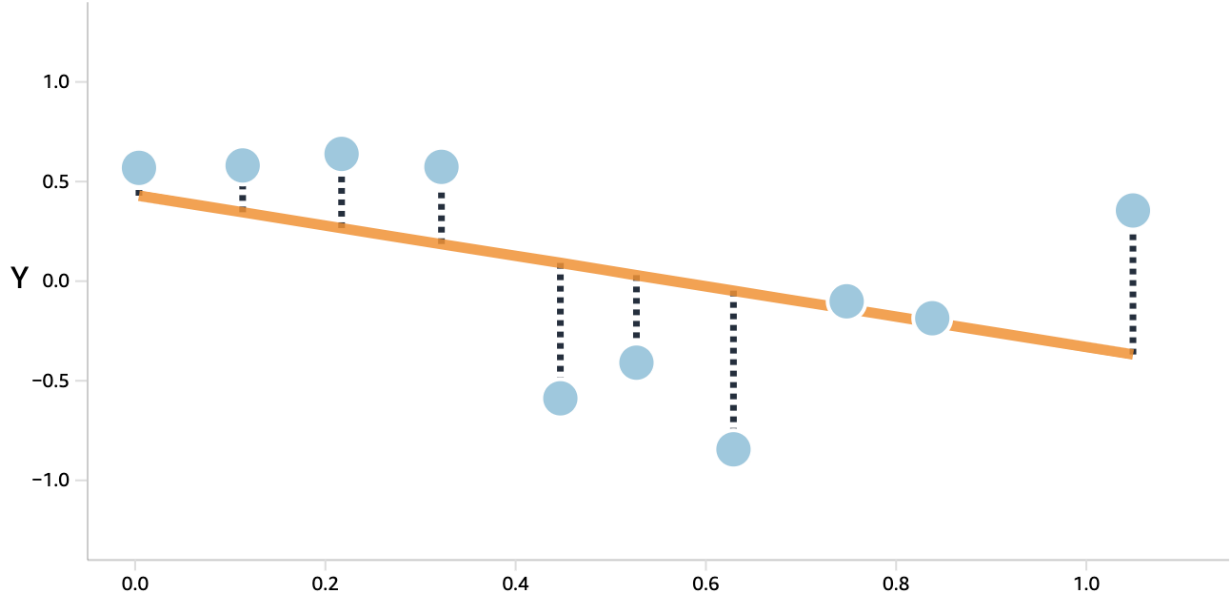
Some lecture slides and instructional materials in this course are adapted from the following sources:

- *An Introduction to Statistical Learning: With Applications in R (Second Edition)*
Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
Springer, 2021

- Online course materials developed by Trevor Hastie, Robert Tibshirani, and collaborators.

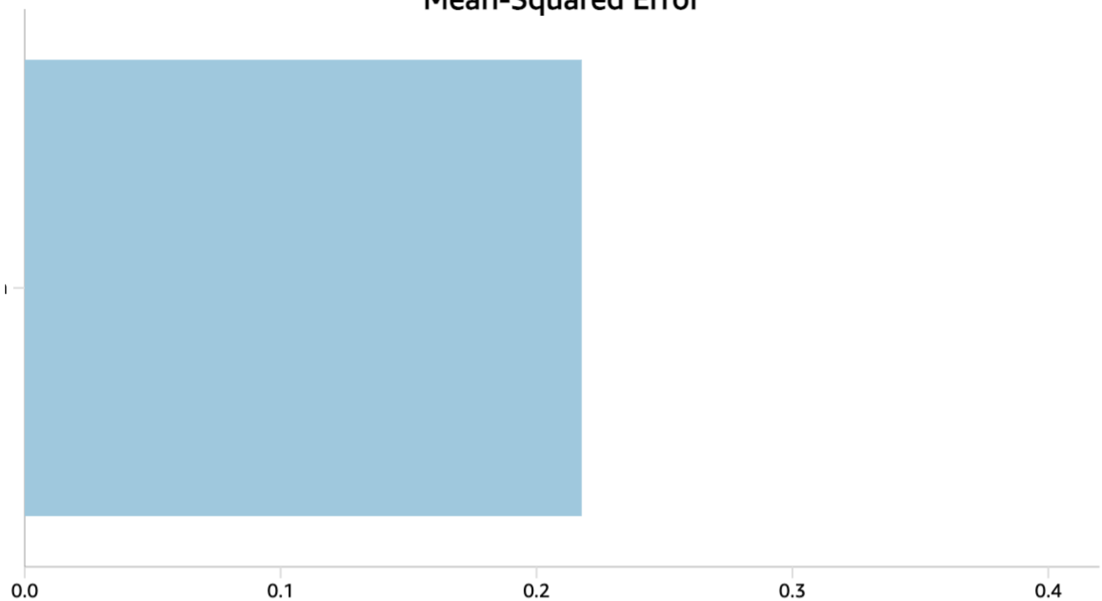


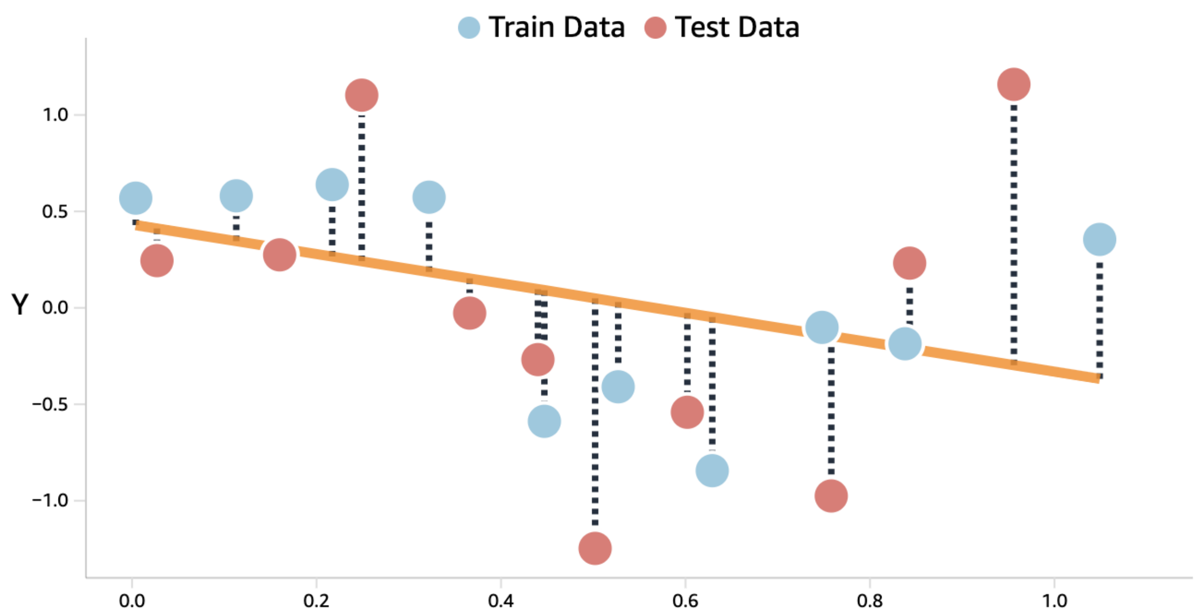
$$\text{Train MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i)^2)$$

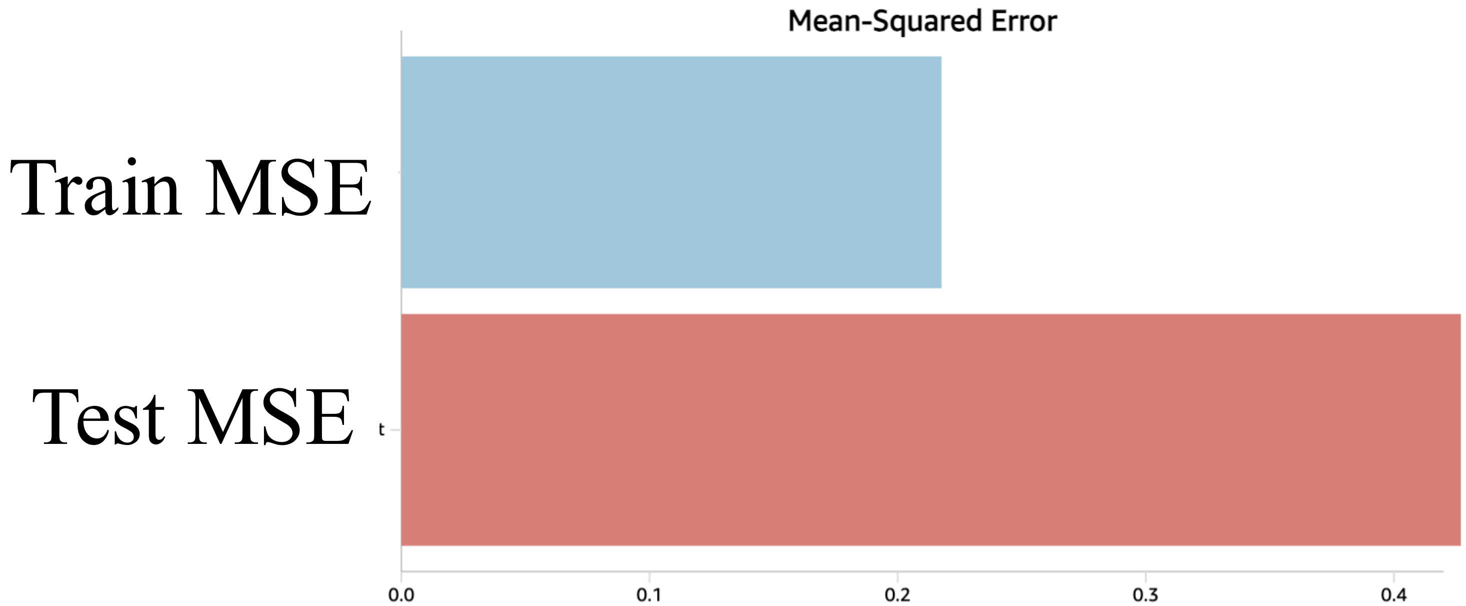


Mean-Squared Error

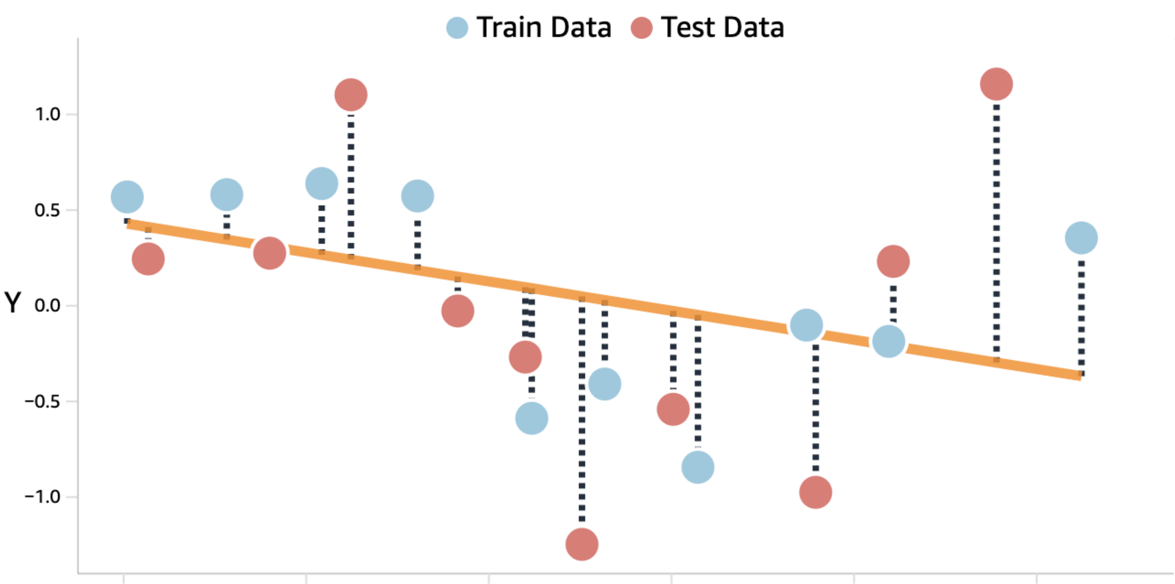
Train MSE

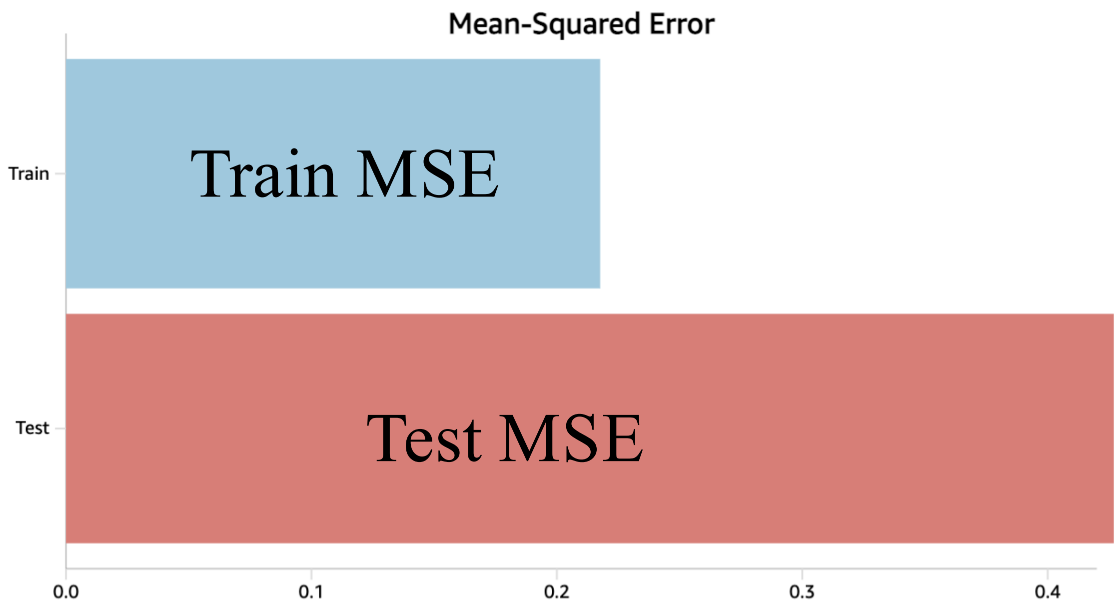



Train Data
Test Data
Y
1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0


Mean-Squared Error
Train MSE
Test MSE
0.0
0.1
0.2
0.3
0.4

LargerTest MSE: Why ?





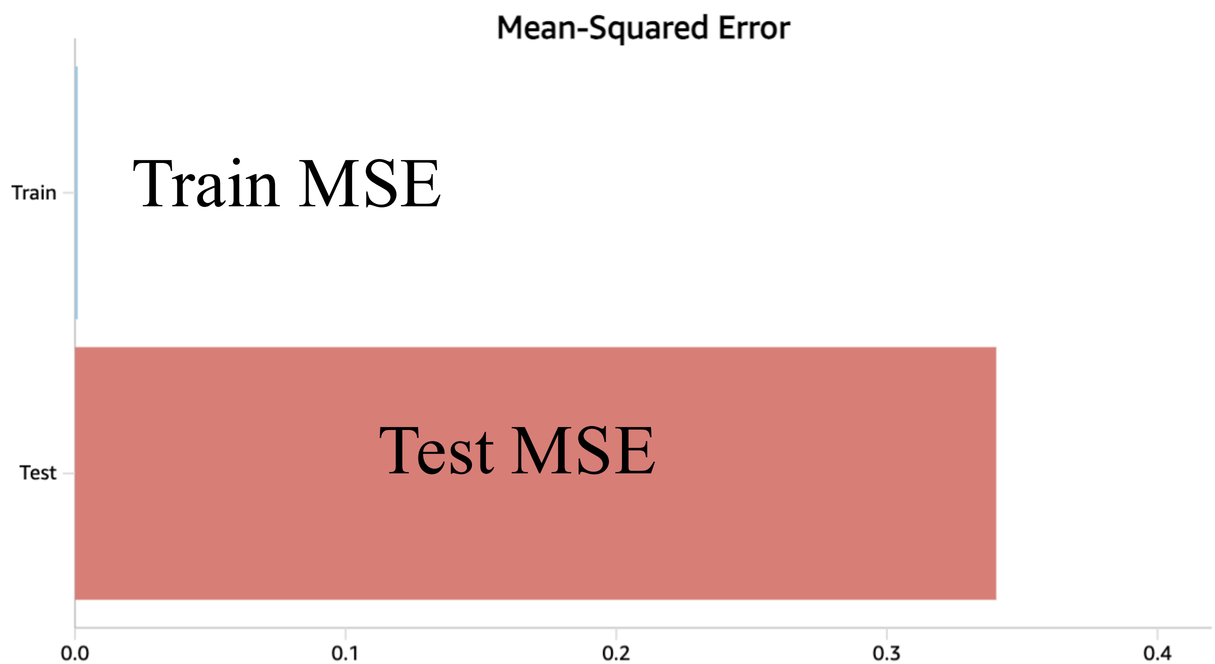
We are using a linear regression model to capture a more complex relationship (our model is too simple)

Underfitting

Large Test MSE: Why ?

Overfitting → models everything including the noise instead of the true relationship

It does NOT generalize well to test data



Bias-variance tradeoff

The test MSE we calculate from our test data is just an estimate for the expected test MSE.

$$\text{test MSE}: \frac{1}{m}\sum_{i=1}^{m}\left(y_i' - \hat{f}(x_i')\right)^2$$

$$\text{expected test MSE}: E\left(y_i' - \hat{f}(x_i')\right)^2$$

Test error composition

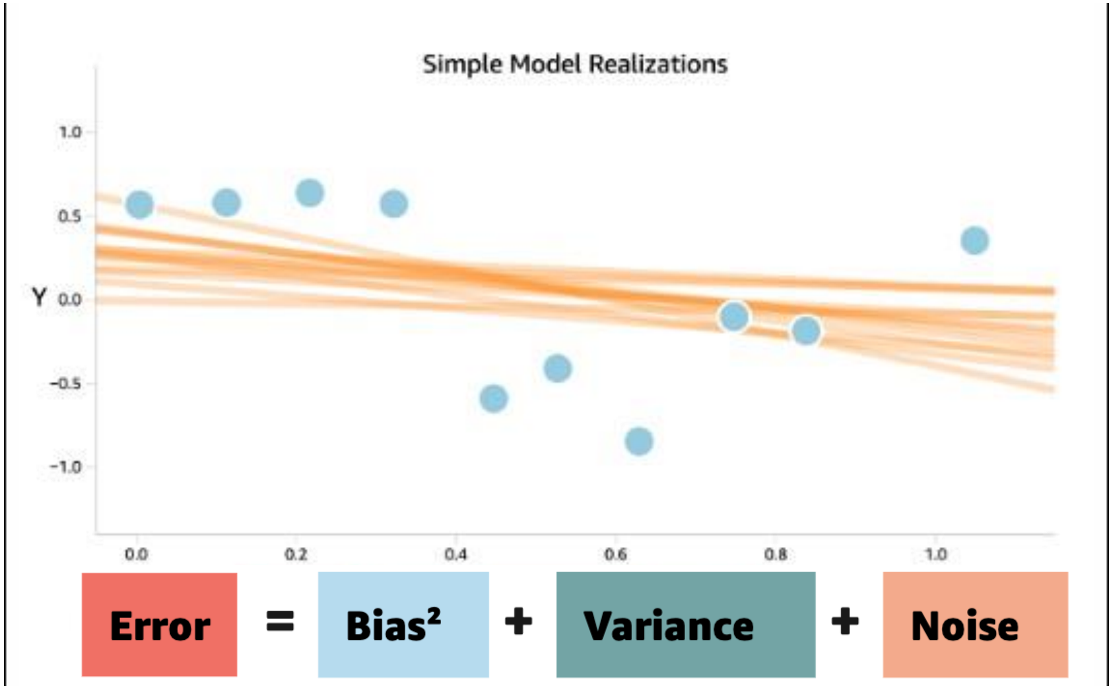
For a fixed value x_0

$$y = f(x_0) + \varepsilon_0$$

$$\mathbb{E}\left(y_0 - \hat{f}(x_0)\right)^2 = \text{Bias}\left(\hat{f}(x_0)\right)^2 + \text{Var}(x_0) + \text{Noise}$$

Bias-variance tradeoff

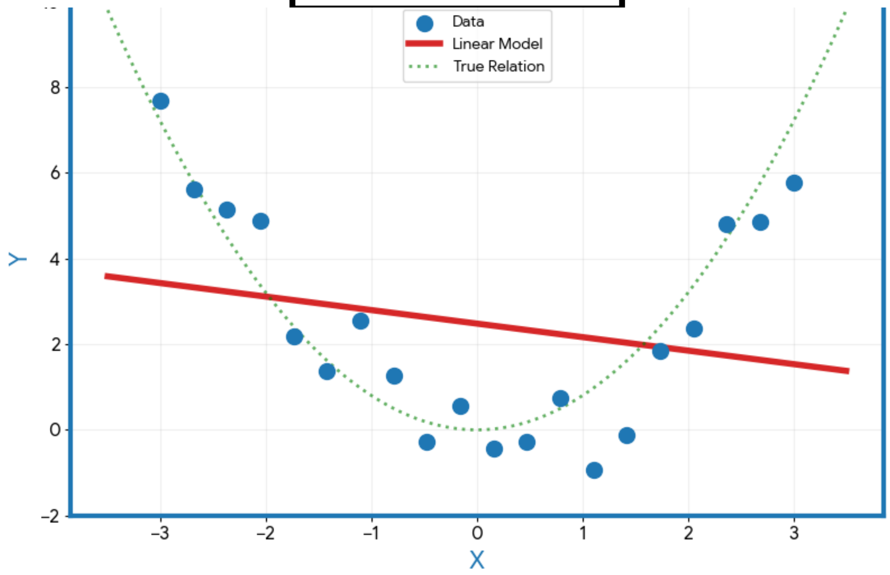
Bias($\hat{f}(x_0)$): refers to the error introduced by estimating f.



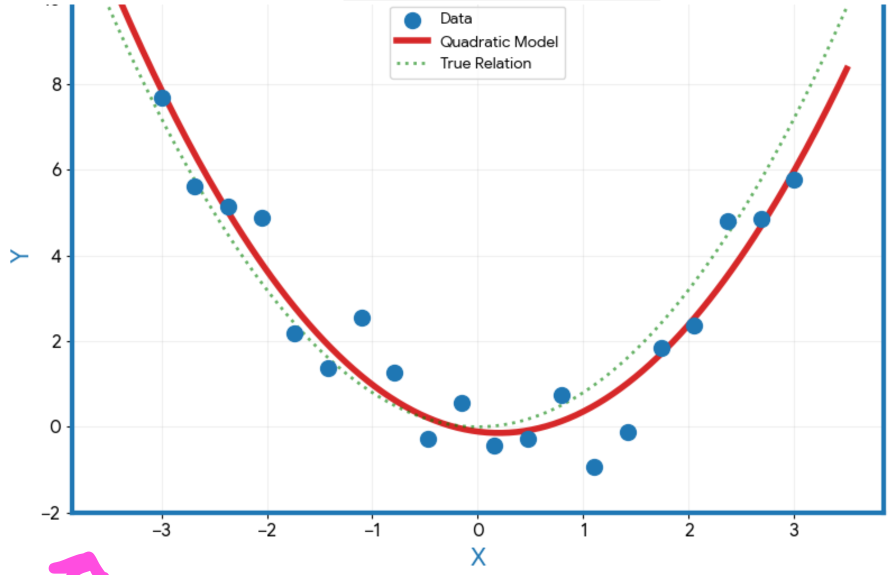
Bias-variance tradeoff

Bias($\hat{f}(x_0)$): refers to the error introduced by estimating f .

High Bias



Low Bias



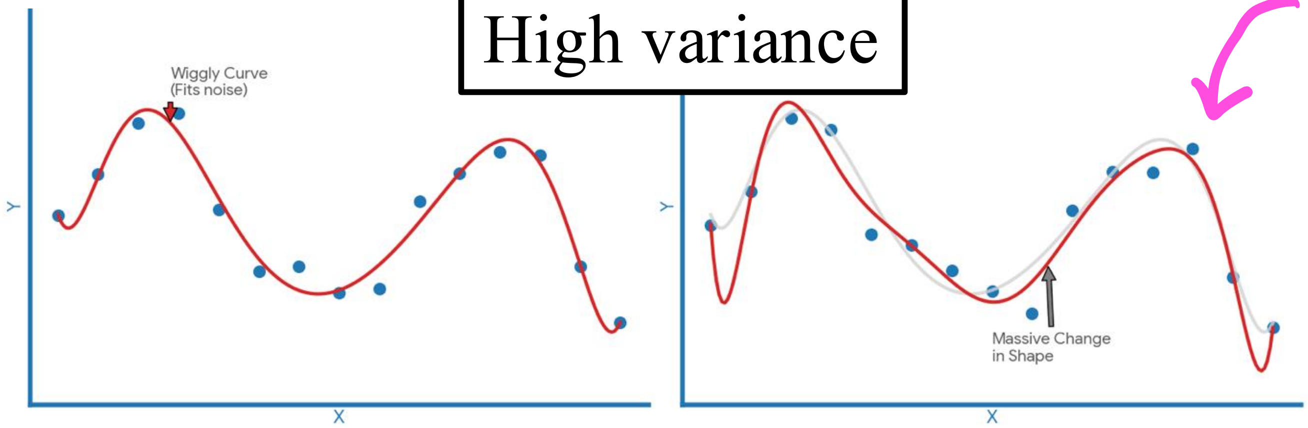
more flexible models tend to have low bias (dif. between true val + pred.)

Bias-variance tradeoff

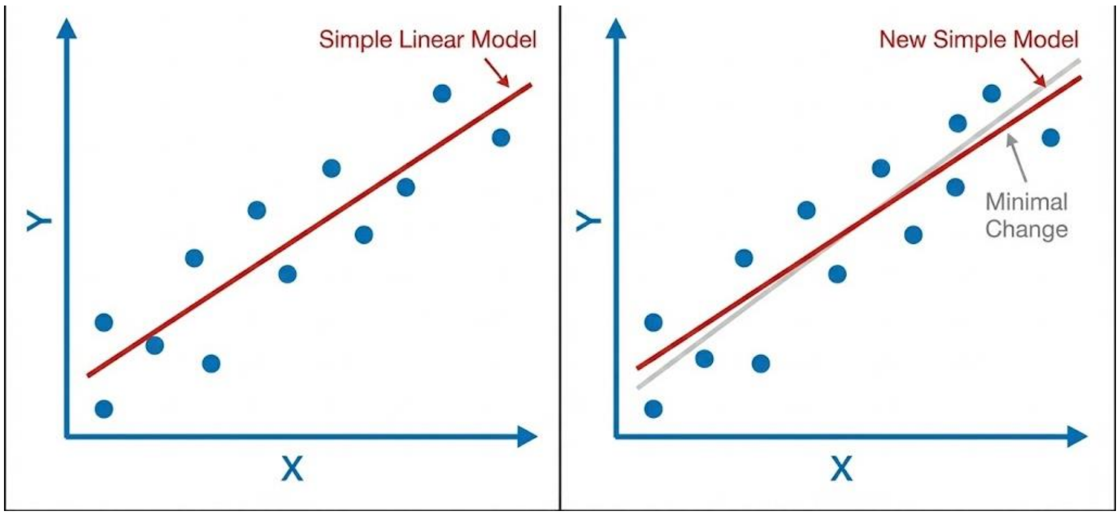
$\mathbf{Var}\left(\hat{f}(x_0)\right)$: the amount by which $\hat{f}(x_0)$ would change if we estimated it using a different training set.

more flexible models tend to have a high variance

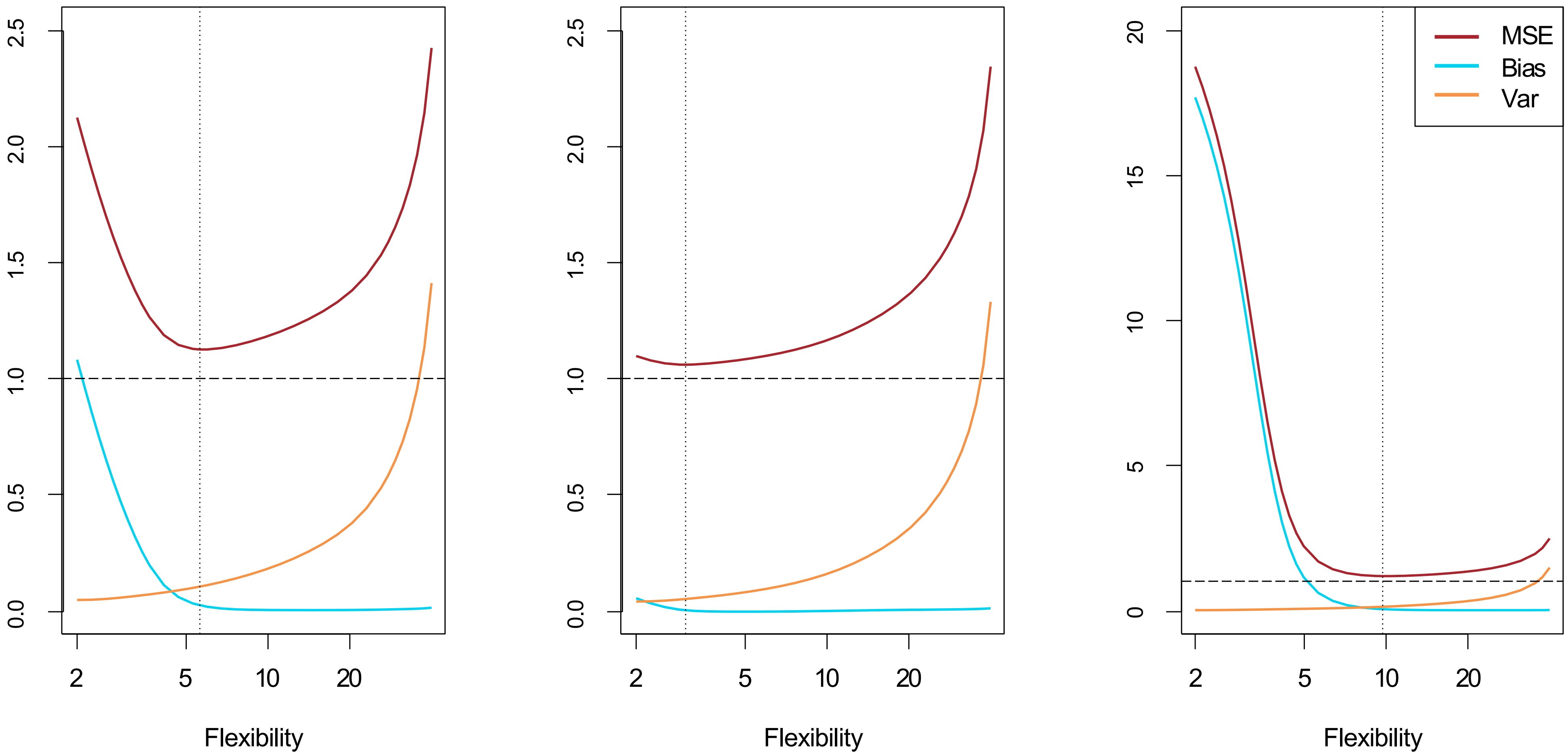
High variance



Low variance

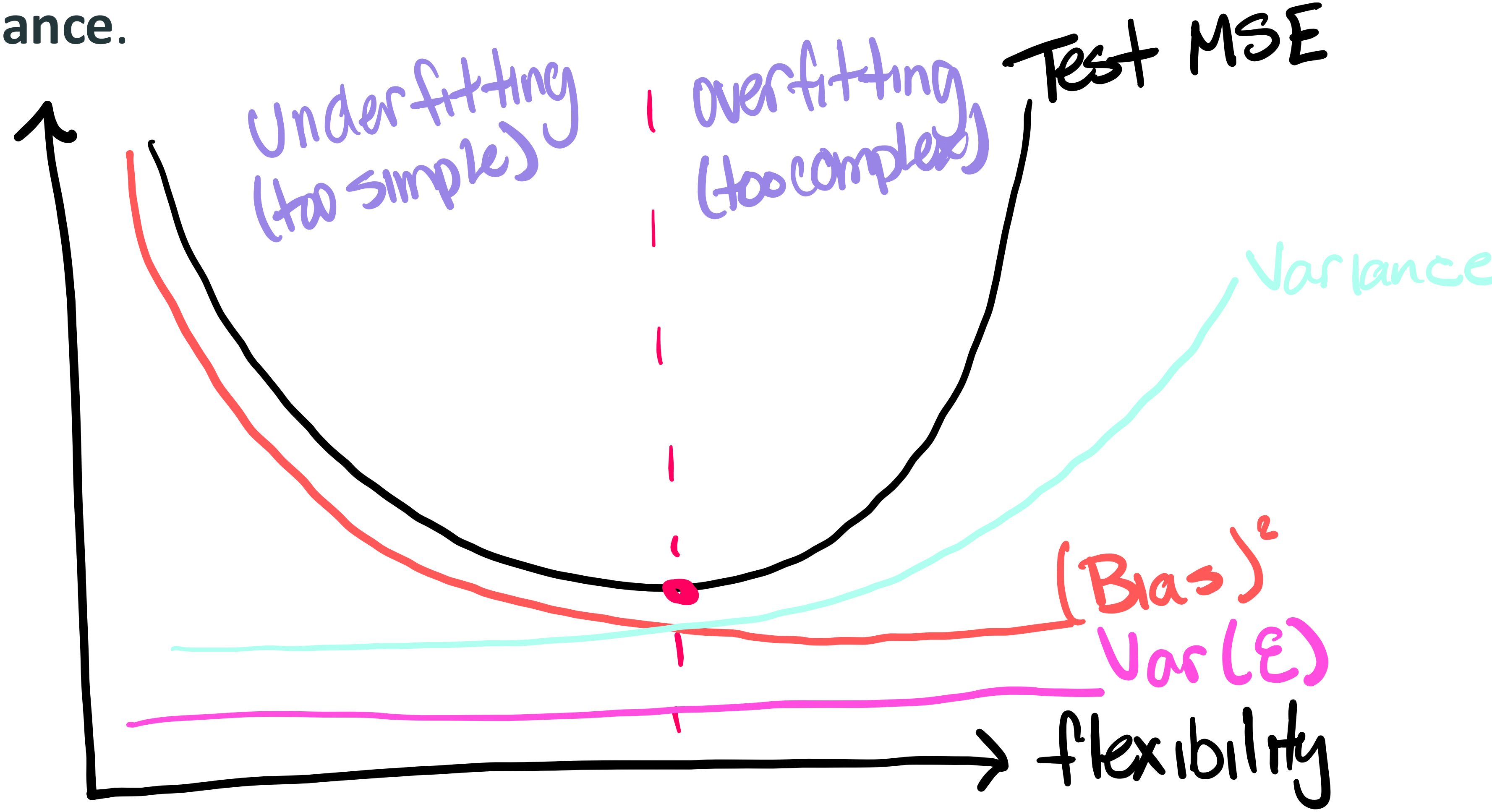


Rate of change depends on the true model



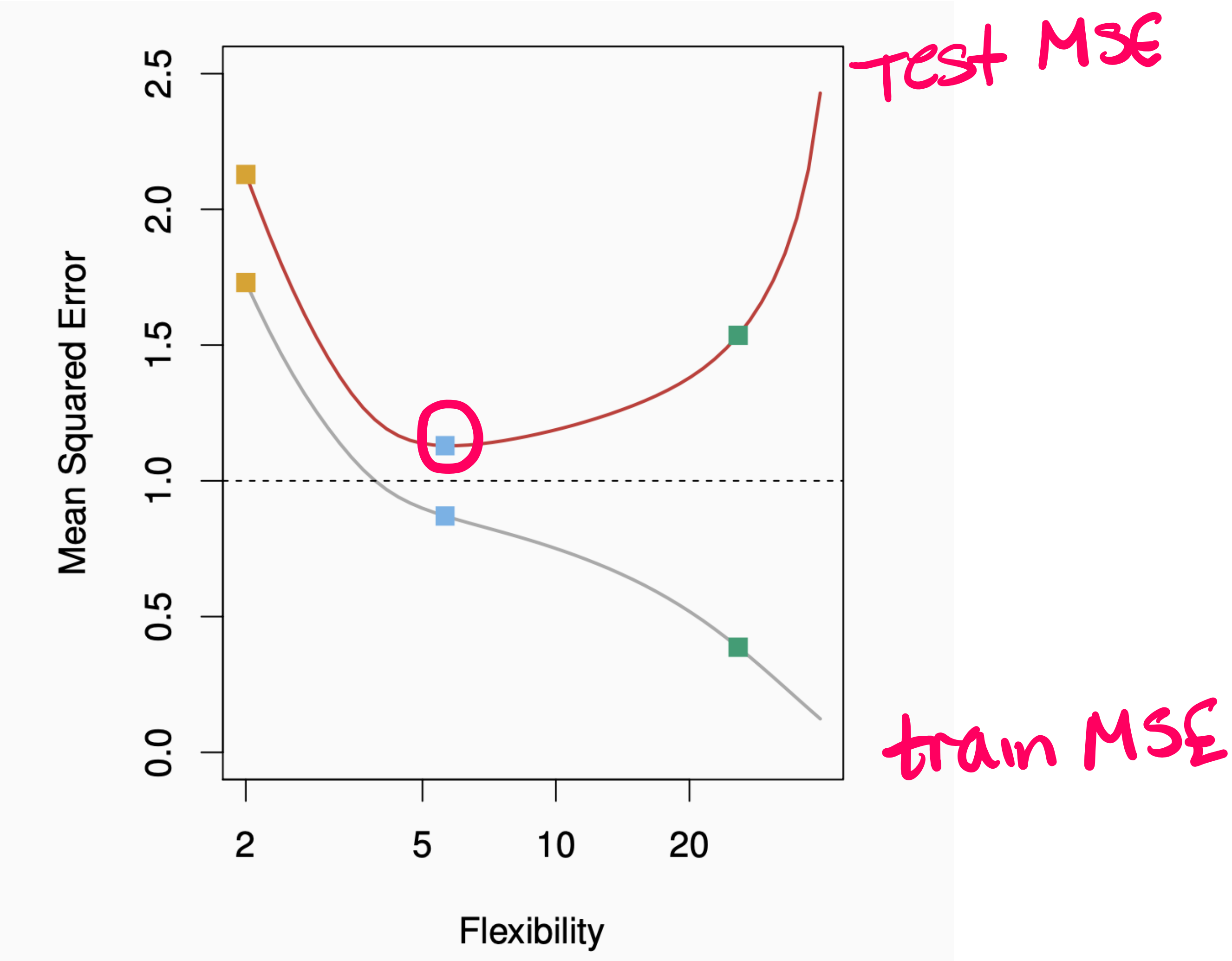
Bias-variance tradeoff

The U-shape observed in the test MSE curves turns out to be the result of two competing properties of statistical learning methods: **bias** and **variance**.



Tradeoff

Our goal in prediction is to select a method that minimizes the test MSE. Low training MSE does not imply low test MSE.



Implications of the bias-variance tradeoff

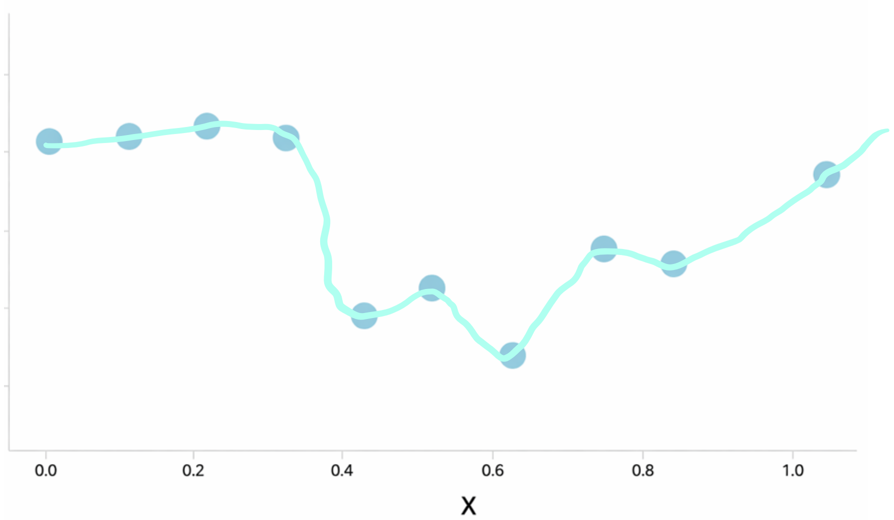
tradeoff

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

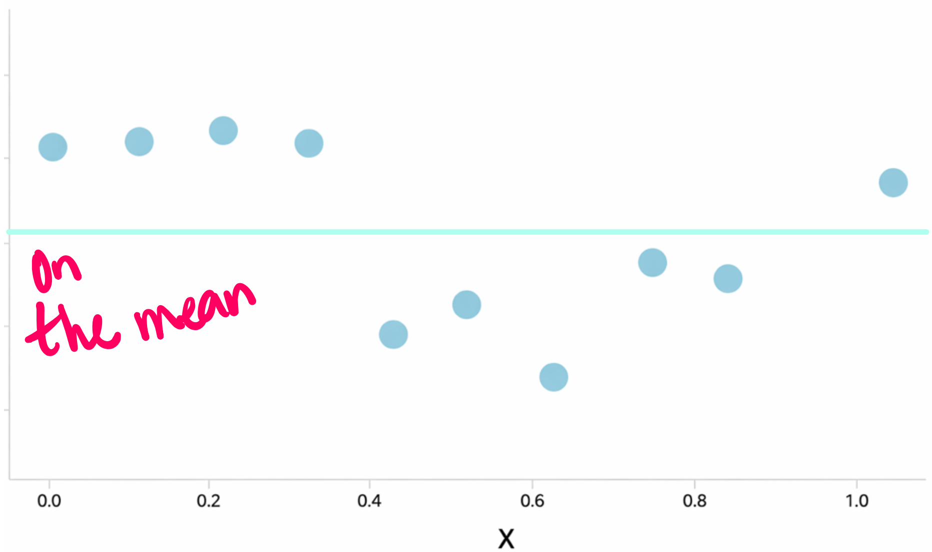
- The expected test MSE is never smaller than the irreducible error.
- Easy to find a zero variance estimate with high bias.
- Easy to find a low bias estimate with high variance.
- In practice, the best expected test MSE is achieved by allowing some bias to decrease variance and vice-versa: this is the bias-variance trade-off.
- General rule: More flexible methods ⇒ higher variance and lower bias.

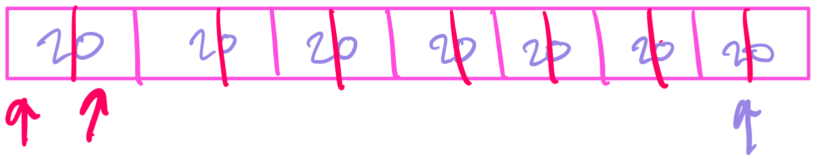
Discussion

(1) Plot a model with zero bias estimate with high variance



(2) Plot a model with a zero variance estimate with high bias.




20
20
20
20
20
20
20