



# Bias-variance Trade-off

Dr. Pei Xu

Auburn University

Tuesday, February 5, 2019

# Measuring Quality of Fit

- One common measure of accuracy for a regression problem is the mean squared error (MSE) i.e.

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

- Where  $\hat{y}_i$  is the prediction our method gives for the observation in our training data.

# Training vs. Test MSE's

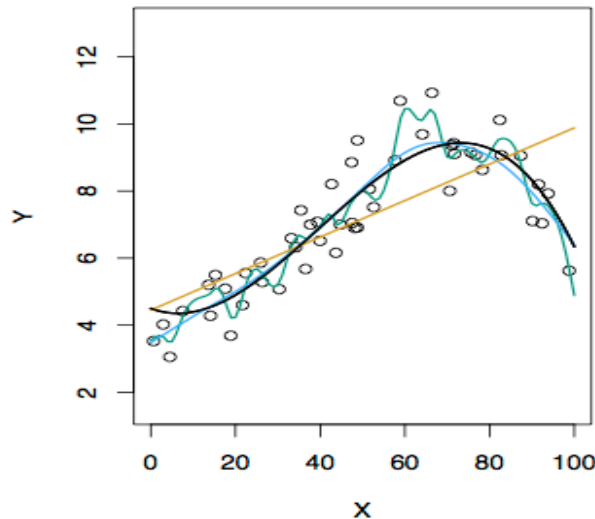
- We want to choose the method that gives the lowest test MSE, as opposed to the lowest training MSE
  - We are interested in the accuracy of the predictions that we obtain when we apply our method to test dataset.
  - Examples:
    - Predict a stock's price
    - Predict risk of diabetes

# Training vs. Test MSE's

- In general the more flexible a method is the lower its training MSE will be i.e. it will “fit” or explain the training data very well.
  - More Flexible methods (such as splines) can generate a wider range of possible shapes to estimate  $f$  as compared to less flexible and more restrictive methods (such as linear regression).
  - The less flexible the method, the easier to interpret the model. Thus, there is a trade-off between flexibility and model interpretability.
- However, the test MSE may in fact be higher for a more flexible method than for a simple approach like linear regression.
- Model Overfitting: is the phenomena where, in the pursuit of minimizing the training error rate, an overly complex model is selected that captures specific patterns in the training data but fails to learn the true nature of relationships in the overall data.

# Examples with Different Levels of Flexibility:

## Example 1



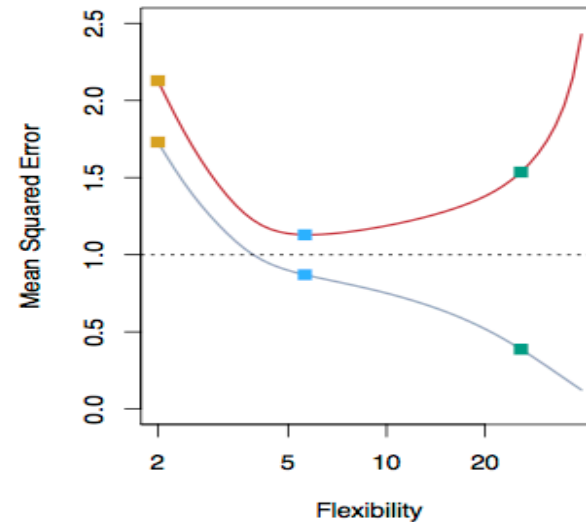
### LEFT

Black: Truth

Orange: Linear Estimate

Blue: smoothing spline

Green: smoothing spline  
(more flexible)



### RIGHT

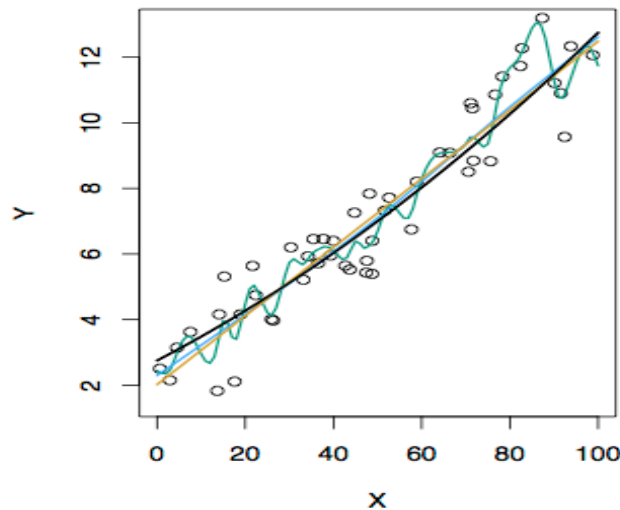
RED: Test MES

Grey: Training MSE

Dashed: Minimum possible  
test MSE (irreducible error)

# Examples with Different Levels of Flexibility:

## Example 2



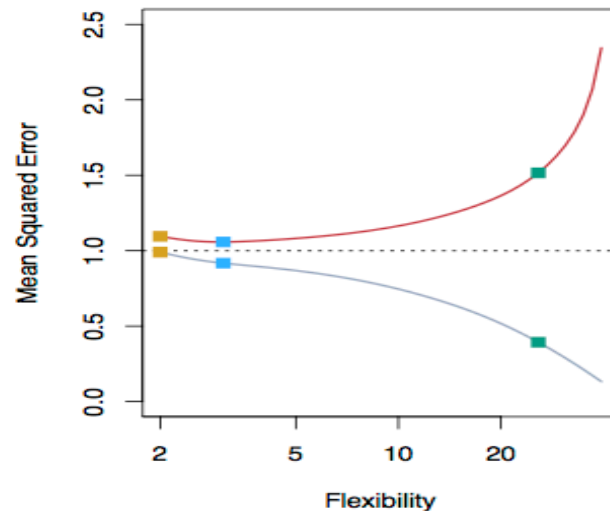
### LEFT

Black: Truth

Orange: Linear Estimate

Blue: smoothing spline

Green: smoothing spline  
(more flexible)



### RIGHT

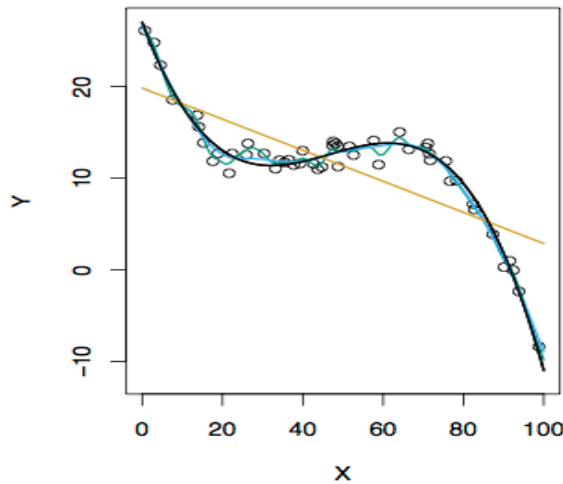
RED: Test MES

Grey: Training MSE

Dashed: Minimum possible  
test MSE (irreducible error)

# Examples with Different Levels of Flexibility:

## Example 3



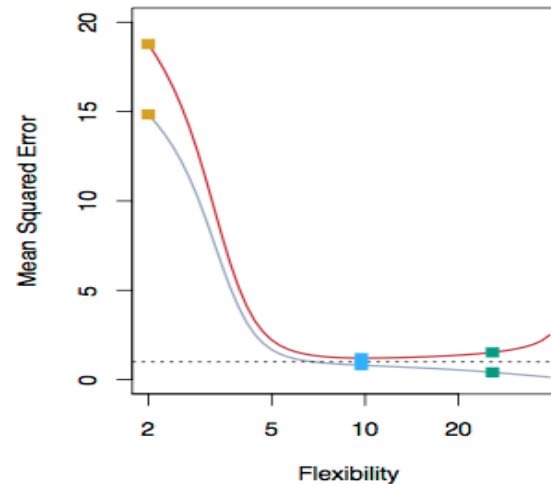
LEFT

Black: Truth

Orange: Linear Estimate

Blue: smoothing spline

Green: smoothing spline  
(more flexible)



RIGHT

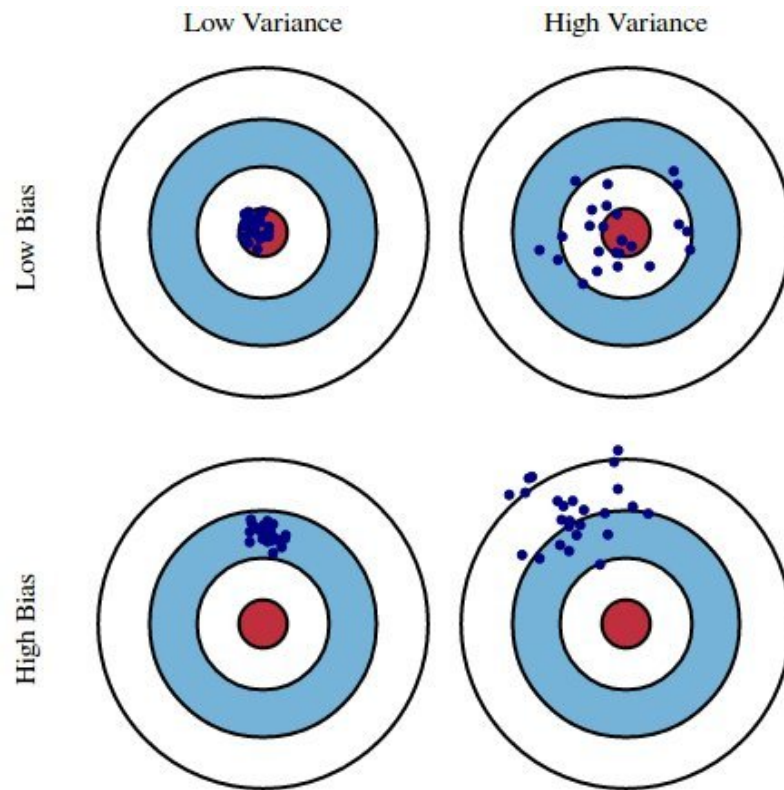
RED: Test MES

Grey: Training MSE

Dashed: Minimum possible  
test MSE (irreducible error)

# Bias/ Variance Tradeoff

- The U-shape observed in the test MSE curves illustrates a very important tradeoff that governs the choice of statistical learning methods.
- There are always two competing forces that govern the choice of learning method i.e. bias and variance.



*Fig. 1: Graphical illustration of bias and variance*



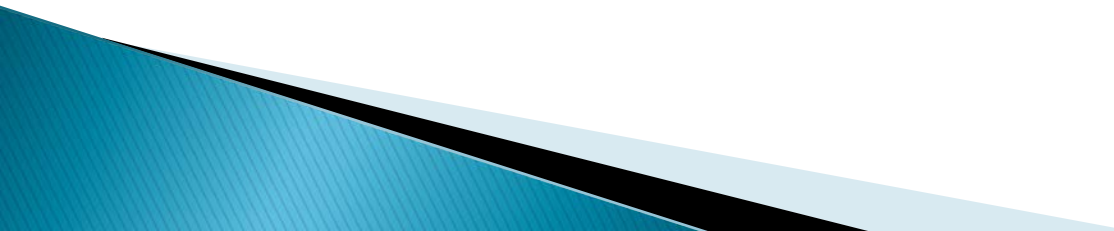
# The Trade-off

- It can be shown that for any given,  $X=x_0$ , the expected test MSE for a new  $Y$  at  $x_0$  will be equal to

$$\text{Expected Test MSE} = E(Y - f(x_0))^2 = \text{Bias}^2 + \text{Var} + \underbrace{\sigma^2}_{\text{Irreducible Error}}$$

- What this means is that as a method gets more complex the bias will decrease and the variance will increase but expected test MSE may go up or down!

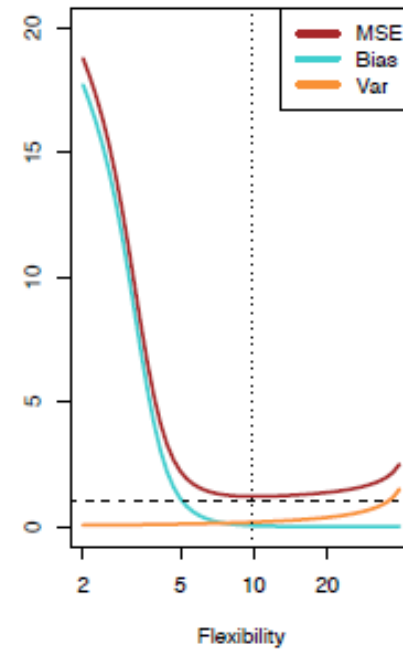
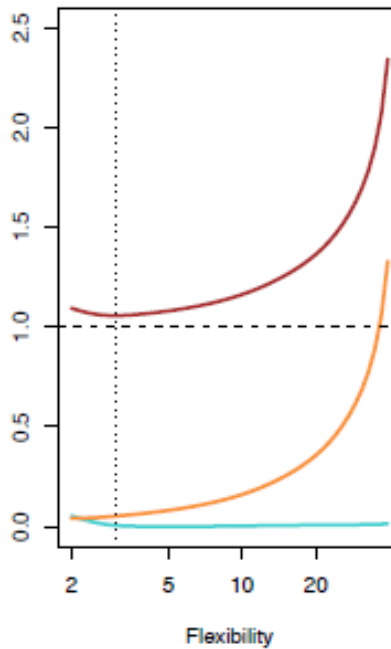
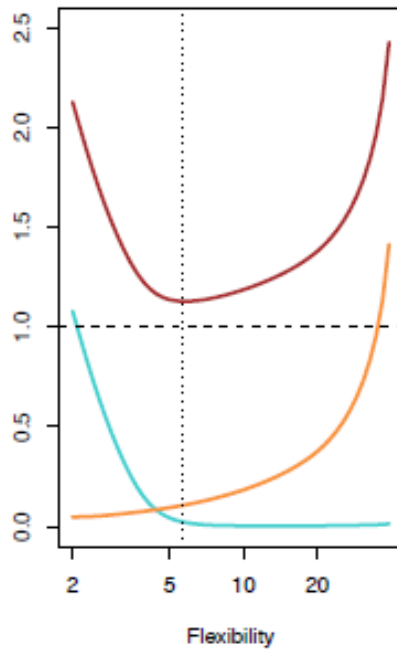
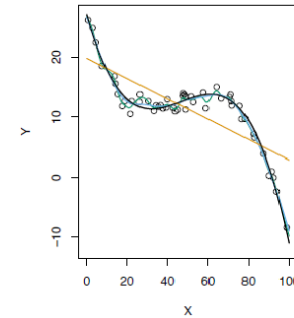
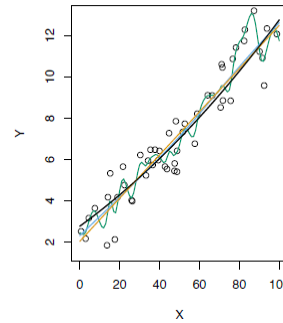
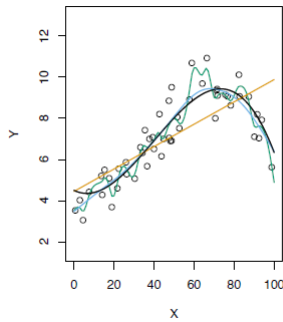
# Bias of Learning Methods

- Bias refers to the error that is introduced by modeling a real life problem (that is usually extremely complicated) by a much simpler problem.
  - For example, linear regression assumes that there is a linear relationship between  $Y$  and  $X$ . It is unlikely that, in real life, the relationship is exactly linear so some bias will be present.
  - The more flexible/complex a method is the less bias it will generally have.
- 

# Variance of Learning Methods

- Variance refers to how much your estimate for  $f$  would change by if you had a different training data set.
- Generally, the more flexible a method is the more variance it has.

# Test MSE, Bias and Variance



# Reference

- ▶ Tan, Pang-Ning et al. Introduction to Data Mining. 2018 (Section 3.4: Model Overfitting)
- ▶ James, Witten, Hastie, and Tibshirani. “An Introduction to Statistical Learning with Applications in R. (Springer; 1st ed. 2013)” ISBN-13: 978-1461471370; ISBN-10: 1461471370 (Chapter 2)