# Ch 2.2: Assessing Model Accuracy

Lecture 3 - CMSE 381

Prof. Lianzhang Bao

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Fri, Jan 17, 2025

# Announcements

Last time:

- Ch 2.1, Vocab day!

Announcements:

- Get on slack!
  - +1 point on the first homework if you post a gif in the thread
- First homework due Sunday, 1/26. Covers:
  - Mon 1/13 lecture
  - Weds 1/15 Lecture
  - Today 1/17 Lecture
- Office hours
  - Tu-W, Dr. Bao (EGR 2507L ) 9am - 10 am
  - Wednesdays, Christy (EGR TBD) 2:30pm - 4:00pm
  - Thursdays, Christy (EGR TBD) 8:45am - 10:15am

# Covered in this lecture

- Mean Squared Error (regression)
- Train vs Test
- Bias Variance Trade off

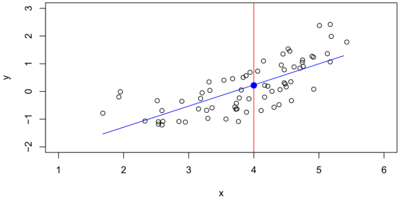
# Quick review of notation

Section 1

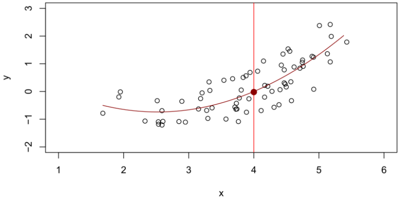
# Mean Squared Error

# Which is better?

A linear model $\hat{f}_L(X) = \hat{\beta}_0 + \hat{\beta}_1 X$ gives a reasonable fit here




A quadratic model $\hat{f}_Q(X) = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2$ fits slightly better.

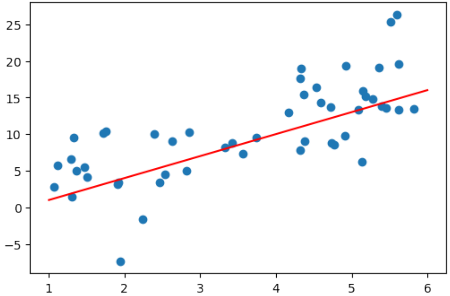


# No free lunch

# Mean Squared Error

## Error in the regression setting

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



# Group Work

Given the following data, you decide to use the model

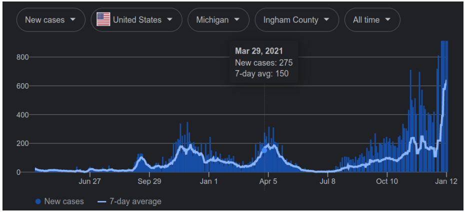
$$\hat{f}(X_1, X_2) = 1 - 3X_1 + 2X_2.$$

What is the MSE?

| X_1 | X_2 | Y |
|---|---|---|
| 0 | 7 | 14 |
| 1 | -3 | -6 |
| 5 | 2 | -10 |
| -1 | 1 | 7 |

1/4 * (1 + 4 + 0 + 1) = 6

# Training MSE



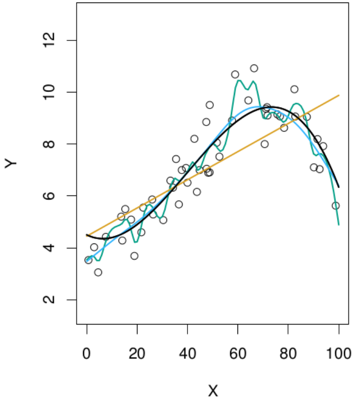
# Train vs test

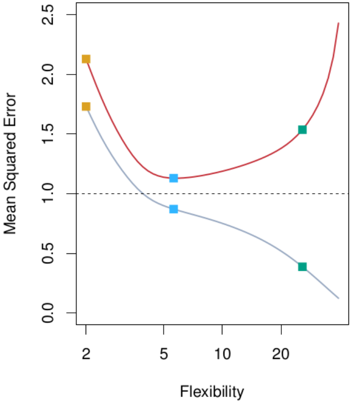
**Training set:**
The observations $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ used to get the estimate $\hat{f}$

**Test set:**
The observations $\{(x'_1, y'_1), \cdots, (x'_{n'}, y'_{n'})\}$ used to compute the average squared prediction error

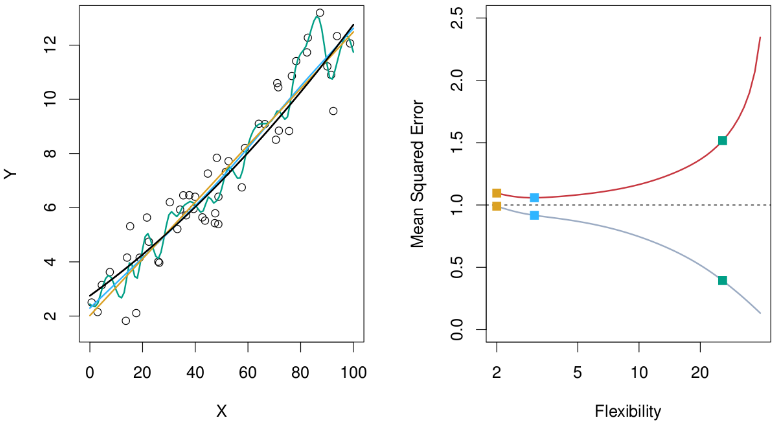
$$\frac{1}{n'} \sum_i (y'_i - \hat{f}(x'_i))^2$$

# Why not just get the best model for the training data?

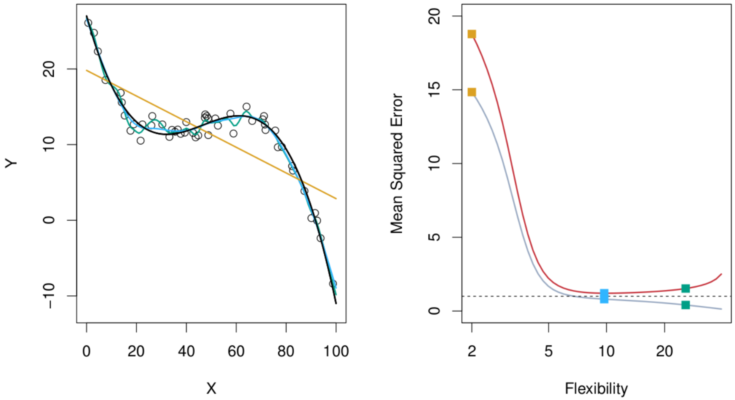


# A more linear example



# A more non-linear example



# A simple solution: Train/test split

## More on this in Ch 5

Section 2

# Bias-Variance Trade-Off

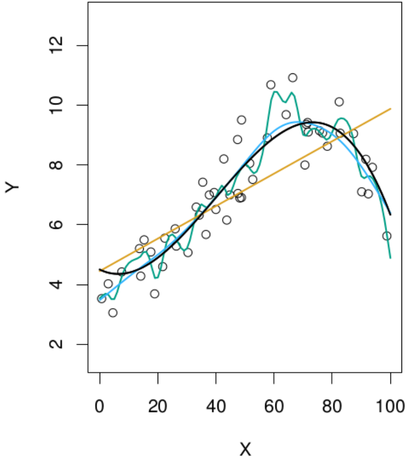
# Bias-variance

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

Reducible

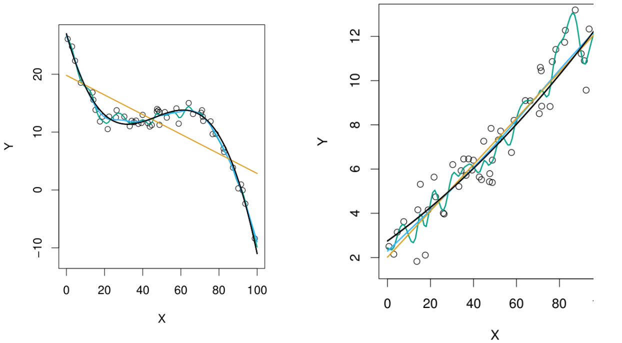
# Variance

**Variance:** the amount by which $\hat{f}$ would change if we estimated it using a different training data set.

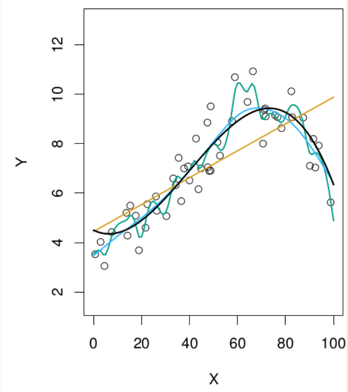


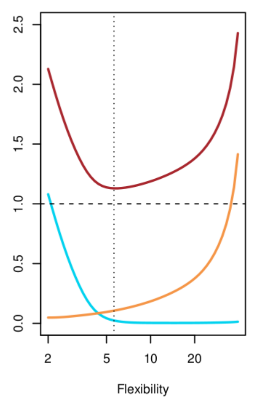
# Bias

**Bias:** the error that is introduced by approximating a (complicated) real-life problem by a much simpler model.



# Group work



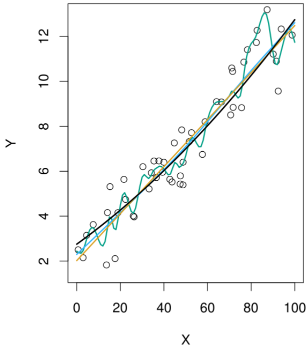



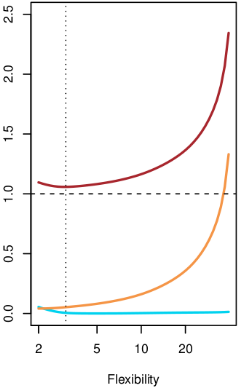

Label the line corresponding to each of the following:

- MSE
- Bias
- Variance of $\hat{f}(x_0)$
- Variance of $\varepsilon$

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

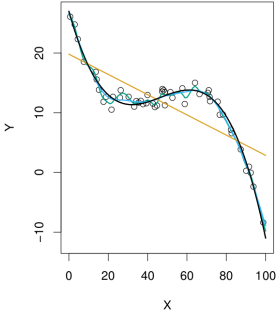
# Another example

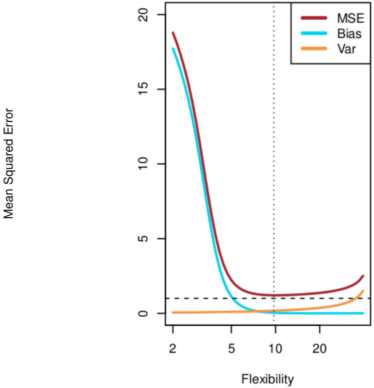



$$E(y_0 - \hat{f}(x_0))^2 = \mathrm{Var}(\hat{f}(x_0)) + \left[\mathrm{Bias}(\hat{f}(x_0))\right]^2 + \mathrm{Var}(\varepsilon)$$

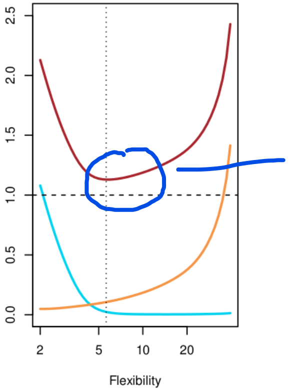
# Yet another example




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

# Bias-variance trade off




ideal

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

# Group work: coding

See jupyter notebook

# Next time

- Wednesday:
  - 3.1 Linear Regression
- Sunday
  - Homework due midnight on D2L

| Lec # | Date | | Topic | Reading | HW | Pop Quizzes | Notes |
|---|---|---|---|---|---|---|---|
| 1 | M | 1/13 | Intro / Python Review | 1 | | | |
| 2 | W | 1/15 | What is statistical learning | 2.1 | | Q1 | |
| 3 | F | 1/17 | Assessing Model Accuracy | 2.2.1, 2.2.2 | | | |
| | M | 1/20 | MLK - No Class | | | | |
| 4 | W | 1/22 | Linear Regression | 3.1 | | Q2 | |
| 5 | F | 1/24 | More Linear Regression | 3.1 | HW #1 Due Sun 1/26 | | |
| 6 | M | 1/27 | Multi-linear Regression | 3.2 | | | |
| 7 | W | 1/29 | Probably More Linear | 3.3 | | Q3 | |