Ch 2.2: Assessing Model Accuracy

Lecture 3 - CMSE 381

Prof. Elizabeth Munch

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

Wed, Sep 4, 2024

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, 9/8. Covers:
  - Mon 8/26 lecture
  - Weds 8/28 Lecture
  - Today 9/4 Lecture
- Office hours
  - Mondays, Dr Munch (Wells TBD) 11am - noon
  - Tuesdays, Christy (EGR TBD) 8:30pm - 10am
  - Thursdays, Christy (EGR TBD) 1 - 2:30pm
  - Fridays, Dr Munch (EGR 1511) 11am - noon

Covered in this lecture

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

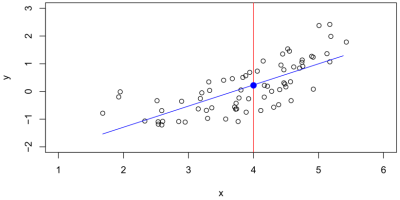
Quick review of notation

Section 1

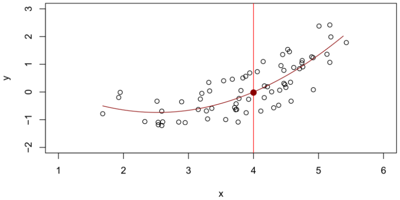
Mean Squared Error

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.

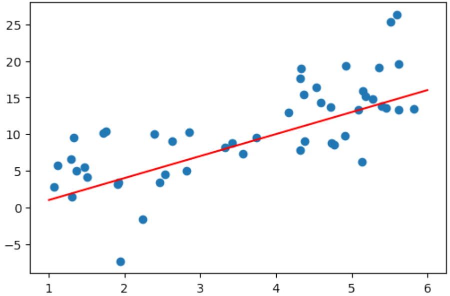


No free lunch

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

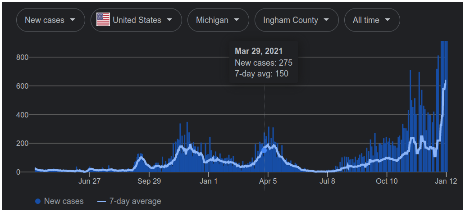
Given the following data, you decide to use the model

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

What is the MSE?

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

Training MSE



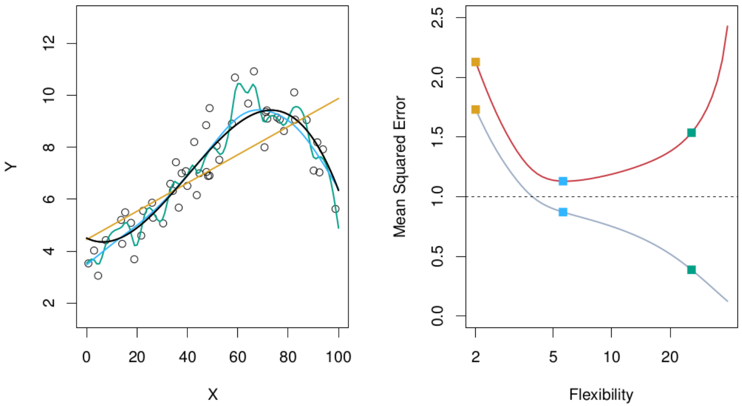
Train vs test

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

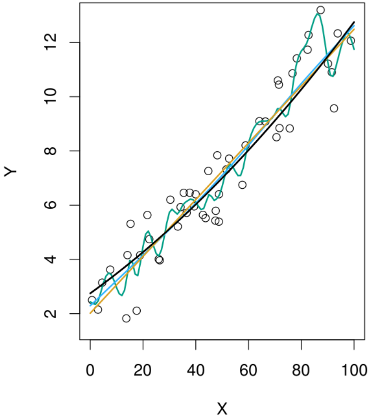
Test set:
The observations $\{(x'_1, y'_1), \cdots, (x'_{n'}, y'_{n'})\}$ used to compute the average squared prediction error

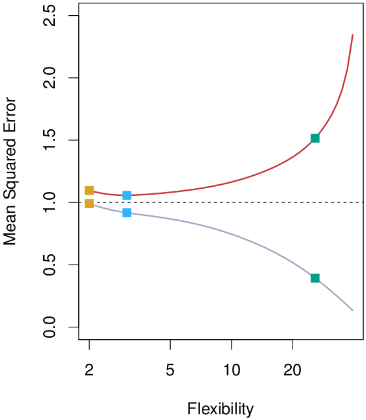
$$\frac{1}{n'} \sum_i (y'_i - \hat{f}(x'_i))^2$$

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

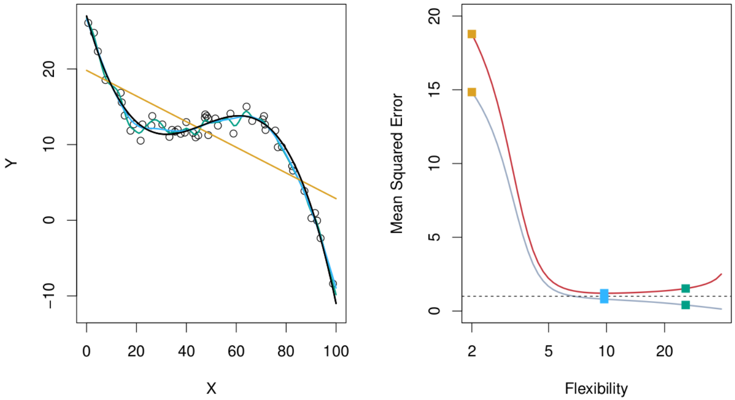


A more linear example





A more non-linear example



A simple solution: Train/test split

More on this in Ch 5

Section 2

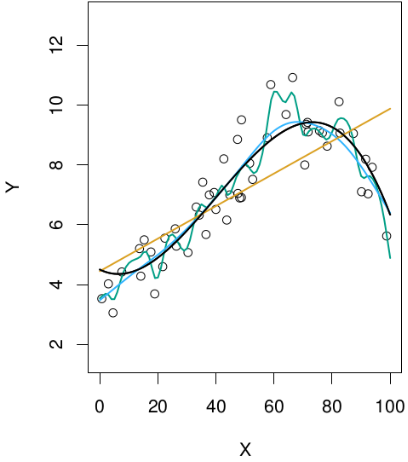
Bias-Variance Trade-Off

Bias-variance

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

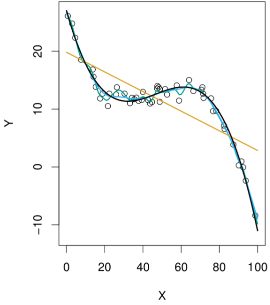
Variance

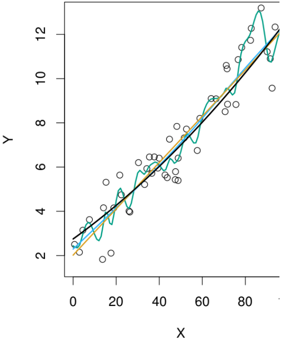
Variance: the amount by which $\hat{f}$ would change if we estimated it using a different training data set.



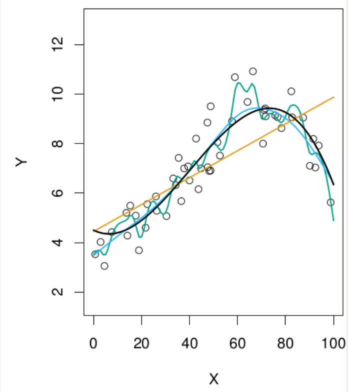
Bias

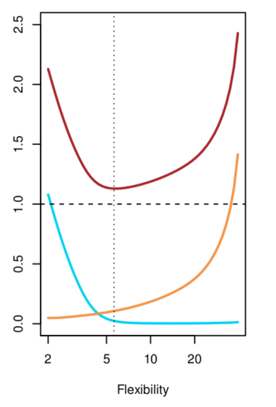
Bias: the error that is introduced by approximating a (complicated) real-life problem by a much simpler model.





Group work



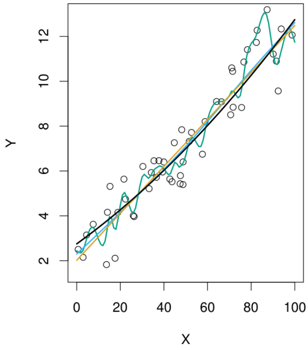


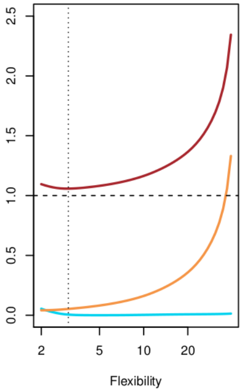
Label the line corresponding to each of the following:

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

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

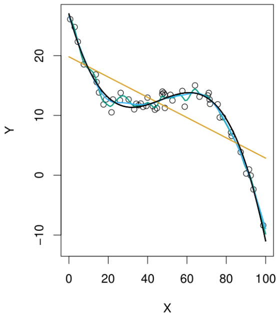
Another example

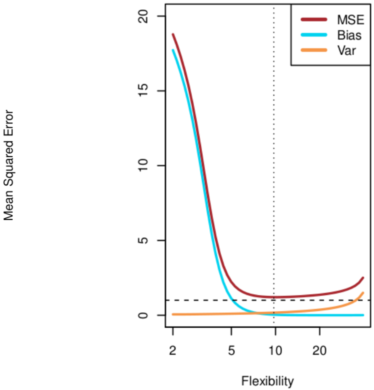




$$E(y_0 - \hat{f}(x_0))^2 = \mathrm{Var}(\hat{f}(x_0)) + \left[\mathrm{Bias}(\hat{f}(x_0))\right]^2 + \mathrm{Var}(\varepsilon)$$

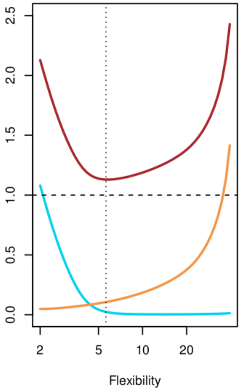
Yet another example





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

Bias-variance trade off



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

Group work: coding

See jupyter notebook

Next time

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

Lec #	Date			Reading	HW
1	Mon	8/26	Intro / First day stuff / Python Review Pt 1	1	
2	Wed	8/28	What is statistical learning?	2.1	
	Fri	8/30	Class Cancelled (Dr Munch out of town)		
	Mon	9/2	No class - Labor day		
3	Wed	9/4	Assessing Model Accuracy	2.2.1, 2.2.2	
4	Fri	9/6	Linear Regression	3.1	HW #1 Due Sun 9/8
5	Mon	9/9	More Linear Regression	3.1/3.2	