Ch 5.1.4-5: More Cross-Validation

Lecture 14 - CMSE 381

Prof. Elizabeth Munch

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

Fri, Oct 6, 2023

Announcements

Last time:

- k-fold CV

This lecture:

- More k-fold CV
- Bias-Variance Tradeoff
- CV for classification

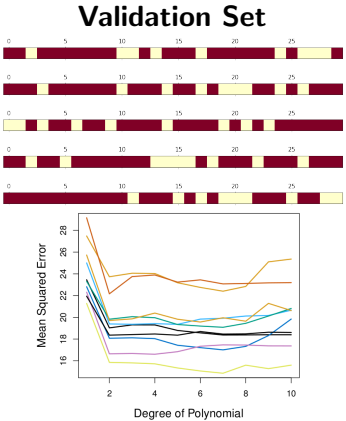
Announcements:

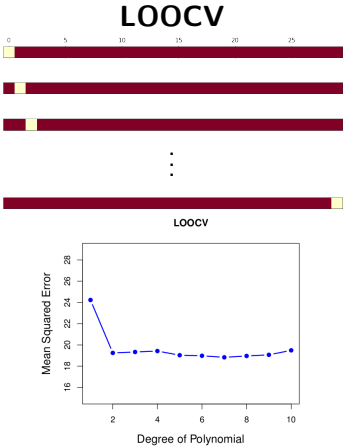
- Homework #4 is posted, Due Monday
-

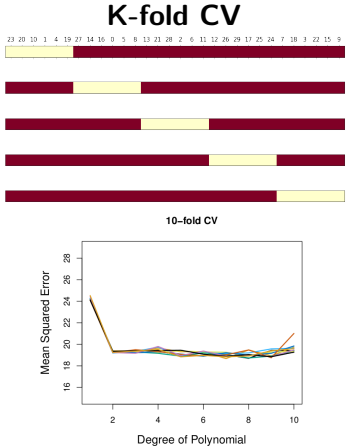
Section 1

k-fold CV

Approximations of Test Error





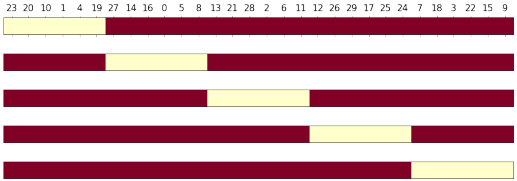


Definition of k-fold CV

- Randomly split data into k-groups (folds)
- Approximately equal sized. For the sake of notation, say each set has ℓ points

- Remove ith fold U_i and reserve for testing.
- Train the model on remaining points
- Calculate $\mathrm{MSE}_i = \frac{1}{\ell}\sum_{(x_j,y_j)\in U_i}(y_j - \hat{y}_j)^2$

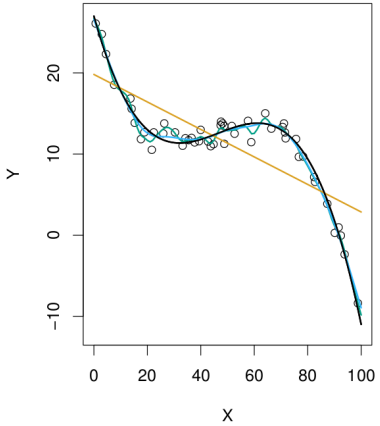
- Rinse and repeat

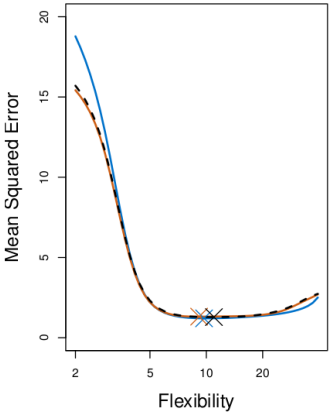


Return

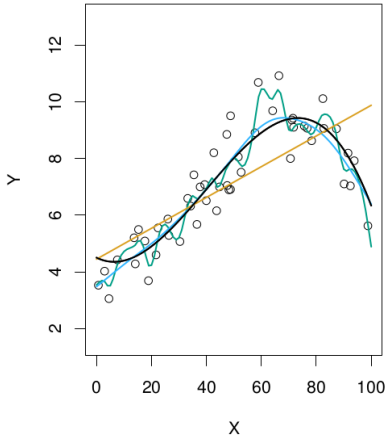
$$CV_{(k)} = \frac{1}{k}\sum_{i=1}^{k}\mathrm{MSE}_i$$

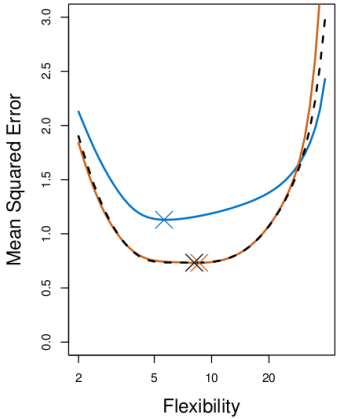
Comparison with simulated data: Ex 3



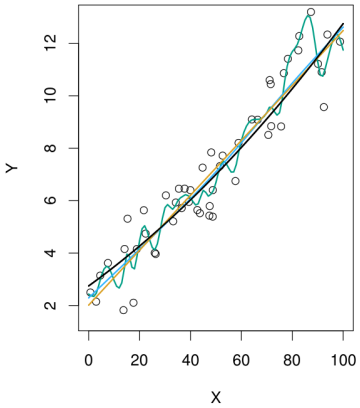


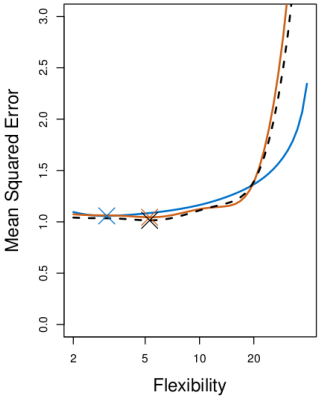
Comparison with simulated data: Ex 1





Comparison with simulated data: Ex 2





Takeaways from the examples

Bias-Variance Tradeoff: Bias

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

Bias-Variance Tradeoff: Variance

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

Bias-Variance Tradeoff

Added frame from later slide deck, might want to incorporate

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

Higher Bias

- Validation set overestimates test error b/c used small subset of data
- k-fold gives medium level of bias b/c training set has approximately $(k-1)n/k$ observations
- LOOCV gives approximately unbiased estimate since uses almost all data every time

Lower Bias

Higher Variance

- LOOCV: avg the outputs of n fitted models with high correlation
- k-fold: k fitted models somewhat less correlated with each other

Lower Variance

Usually use $k = 5$ or $k = 10$

Section 2

Aside - Polynomial linear regression

Polynomial regression

Replace linear model

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$$

with

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \cdots + \beta_d x_1^d + \varepsilon_i$$

Faking linear regression into doing our work for us

Coding - Build a plot for train/test scores vs flexibility

Next time

12	Mon	Oct 2	Leave one out CV	5.1.1, 5.1.2	
13	Wed	Oct 4	k-fold CV	5.1.3	
14	Fri	Oct 6	More k-fold CV,	5.1.4-5	
15	Mon	Oct 9	k-fold CV for classification	5.1.5	HW #4 Due
16	Wed	Oct 11	Resampling methods: Bootstrap	5.2	
17	Fri	Oct 13	Subset selection	6.1	
18	Mon	Oct 16	Shrinkage: Ridge	6.2.1	
19	Wed	Oct 18	Shrinkage: Lasso	6.2.2	
	Fri	Oct 20	**Review**		
	Mon	Oct 23	No class - Fall break		
	Wed	Oct 25	**Midterm #2**		
20	Fri	Oct 27	Dimension Reduction	6.3	