Ch 5.1.3-4: k-Fold Cross-Validation

Lecture 11 - CMSE 381

Prof. Elizabeth Munch

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

Wed, Oct 5, 2022

Announcements

Last time:

- Validation Set
- LOOCV

Announcements:

-
-

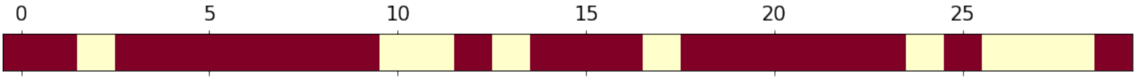
Covered in this lecture

- k-fold CV
- CV for classification

Section 1

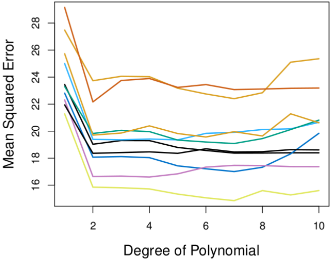
Last time

Validation set approach

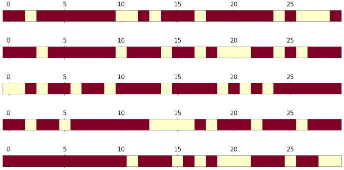


- Divide randomly into two parts:
  - Training set
  - Validation/Hold-out/Testing set
- Fit model on training set
- Use fitted model to predict response for observations in the test set
- Evaluate quality (e.g. MSE)

Problems



Ex. Predict mpg using horsepower



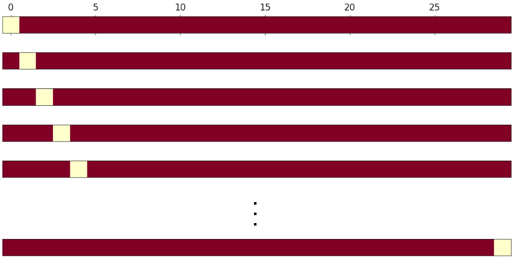
- Highly variable results, no consensus about the error
- Tends to overestimate test error rate

Leave One Out CV (LOOCV)

- Remove (x_1, y_1) for testing.
- Train the model on $n-1$ points: $\{(x_2, y_2), \cdots, (x_n, y_n)\}$
- Calculate $\mathrm{MSE}_1 = (y_1 - \hat{y}_1)^2$

- Remove (x_2, y_2) for testing.
- Train the model on $n-1$ points: $\{(x_1, y_1), (x_3, y_3), \cdots, (x_n, y_n)\}$
- Calculate $\mathrm{MSE}_2 = (y_2 - \hat{y}_2)^2$

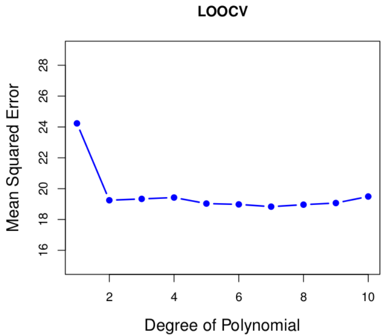
- Rinse and repeat



Return the score:

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

Pros and Cons



- No variance
- Higher computation cost

Speeding up LOOCV

Warning: This only works for least squares linear or polynomial regression.

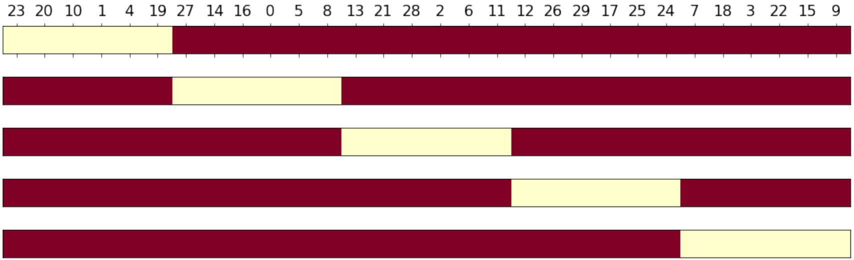
$$h_i = \frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum_{j=1}^{n}(x_j - \overline{x})^2}$$

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

Section 2

k-Fold CV

The idea


23 20 10 1 4 19 27 14 16 0 5 8 13 21 28 2 6 11 12 26 29 17 25 24 7 18 3 22 15 9

Mathy version

- 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

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

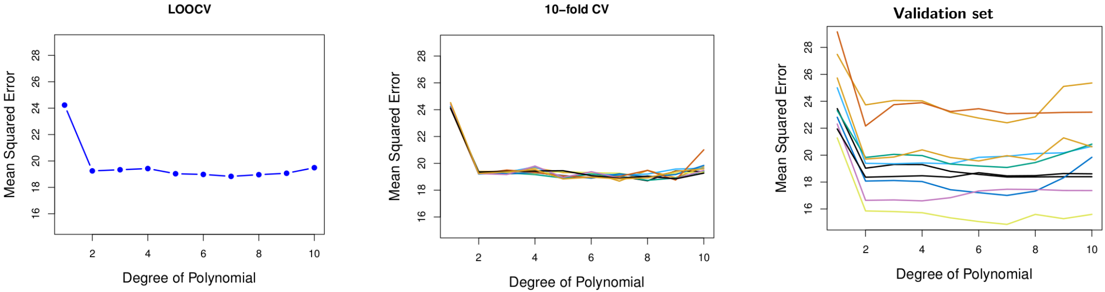
Coding - Building k-fold CV

Pros and Cons

Pros:

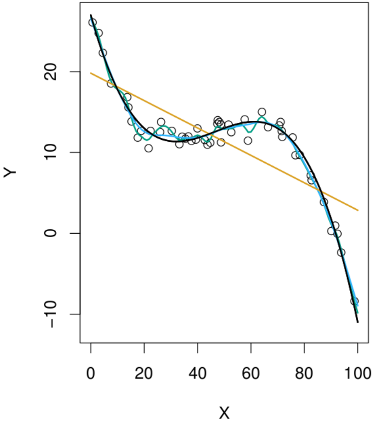
Cons:

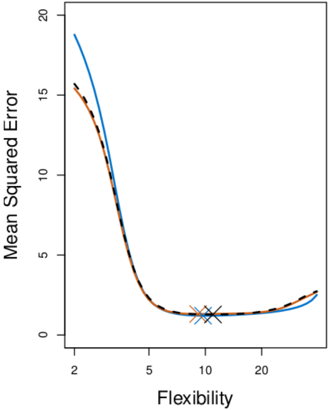
Comparison



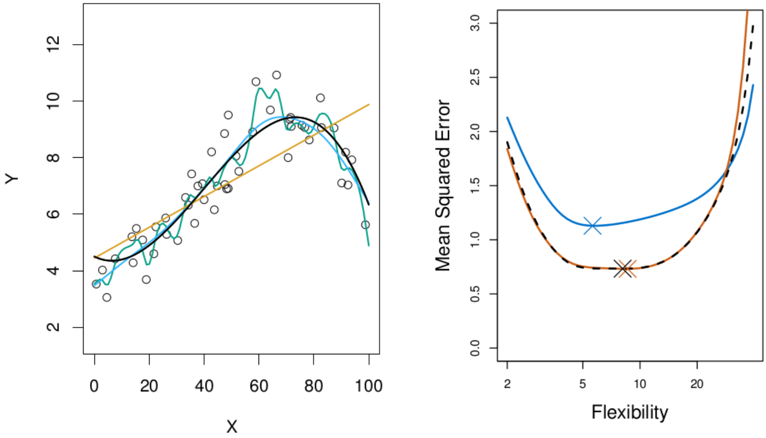
LOOCV
10−fold CV
Validation set
Mean Squared Error
Degree of Polynomial
16
18
20
22
24
26
28
2
4
6
8
10

Comparison with simulated data: Ex 3

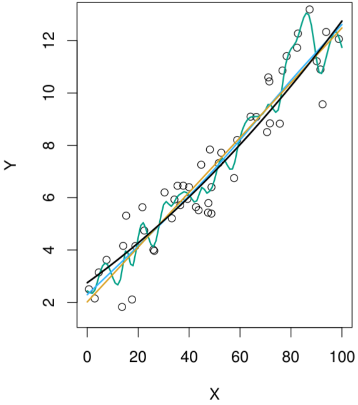


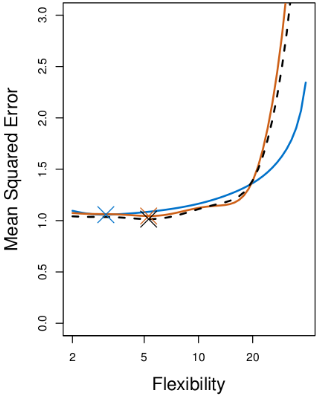


Comparison with simulated data: Ex 1



Comparison with simulated data: Ex 2





Takeaways from the examples

Bias-Variance Tradeoff: Bias

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

Bias-Variance Tradeoff: Variance

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

Do the remainder of the coding

Next time

14	M	Oct 3	Leave one out CV	5.1.1, 5.1.2	
15	W	Oct 5	k-fold CV	5.1.3	
16	F	Oct 7	More k-fold CV	5.1.4	
17	M	Oct 10	CV for classification	5.1.5	HW #4 Due
18	W	Oct 12	Resampling methods: Bootstrap	5.2	
19	F	Oct 14	Subset selection	6.1	
20	M	Oct 17	Shrinkage: Ridge	6.2.1	HW #5 Due
21	W	Oct 19	Shrinkage: Lasso	6.2.2	
22	F	Oct 21	Dimension Reduction	6.3	
	M	Oct 24	No class - Fall break		
21	W	Oct 26	More dimension reduction; High dimensions	6.4	
22	F	Oct 28	Polynomial & Step Functions.	7.1,7.2	HW #6 Due
23	M	Oct 31	Review		
24	W	Nov 2	***Midterm #2***		