Ch 5.1.4-5: More Cross-Validation

Lecture 14 - CMSE 381

Prof. Guanqun Cao

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

Wed, Oct 1, 2025

Announcements

Last time:

- k-fold CV

This lecture:

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

Announcements:

- Exam 1 feedback sent

CMSE381_F2025_Schedule : Schedule

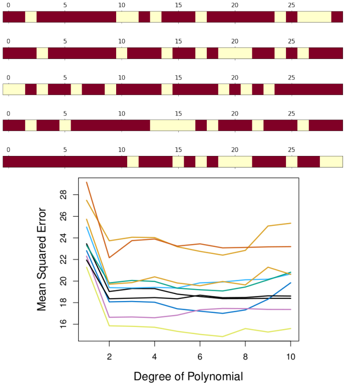
	M	9/22	***Project Day & Review***		
	W	9/24	**Midterm #1**		
12	F	9/26	Leave one out CV	5.1.1, 5.1.2	
13	M	9/29	k-fold CV	5.1.3	
14	W	10/1	More k-fold CV	5.1.4-5	
15	F	10/3	k-fold CV for classification	5.1.5	
16	M	10/6	Subset selection	6.1	
17	W	10/8	Shrinkage: Ridge	6.2.1	
18	F	10/10	Shrinkage: Lasso	6.2.2	HW #4 Due Sun 10/12
19	M	10/13	PCA	6.3	
20	W	10/15	PCR	6.3	
	F	10/17	***Review***		
	M	10/20	Fall Break		
	W	10/22	**Midterm #2**		
21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/28
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	W	10/29	Regression Splines	7.4	

Section 1

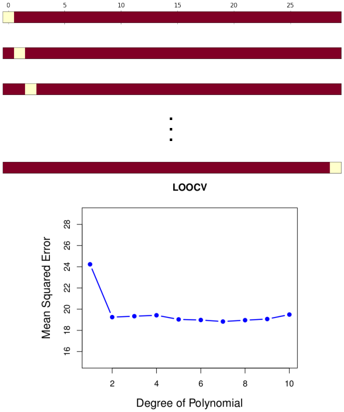
k-fold CV

Approximations of Test Error

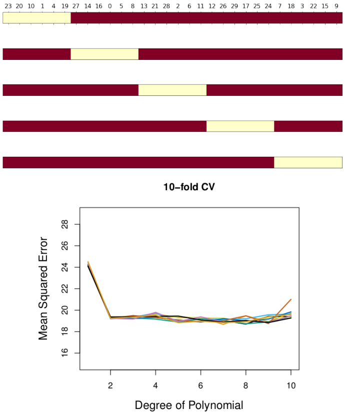
Validation Set



LOOCV

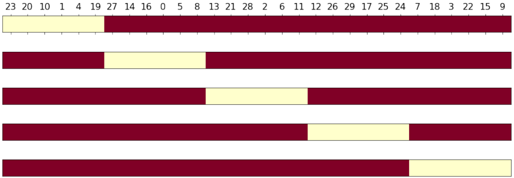


K-fold CV



Definition of k-fold CV

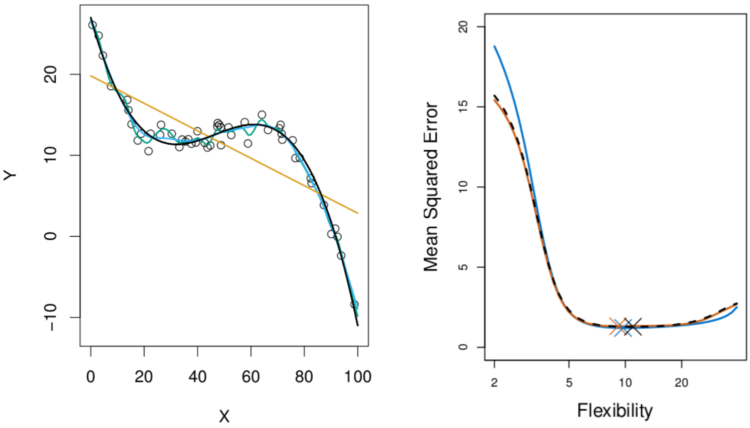
- Randomly split data into k-groups (folds)
- Approximately equal sized. For the sake of notation, say each set has ℓ points
- Remove ith fold U_i and reserve for testing.
- Train the model on remaining points
- Calculate $\mathrm{MSE}_i = \frac{1}{\ell}\sum_{(x_j,y_j)\in U_i}(y_j - \hat{y}_j)^2$
- Rinse and repeat



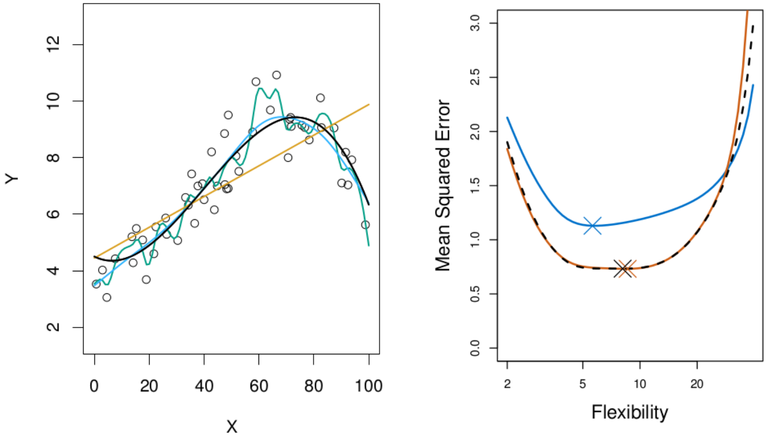
Return

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

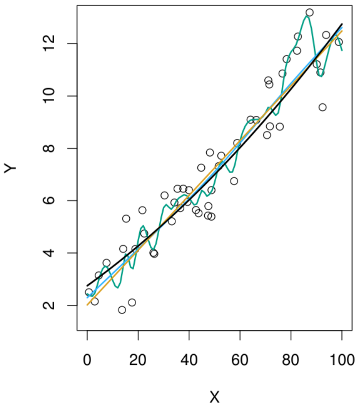
Comparison with simulated data: Ex 3

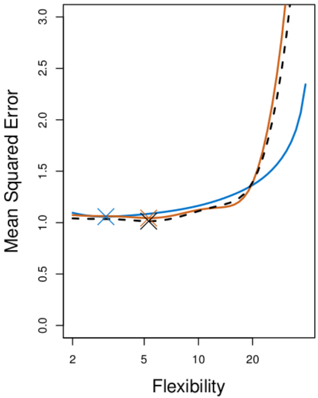


Comparison with simulated data: Ex 1



Comparison with simulated data: Ex 2





Takeaways from the examples

Bias-Variance Tradeoff: Bias

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

Bias-Variance Tradeoff: Variance

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

In short: Vadidation vs Test

- all the time, we are pretending the validation set etc is the test set...
- when it is not.

Real-world example: Chekroud et al., Science 383, 164–167 (2024)

RESEARCH

RESEARCH ARTICLE

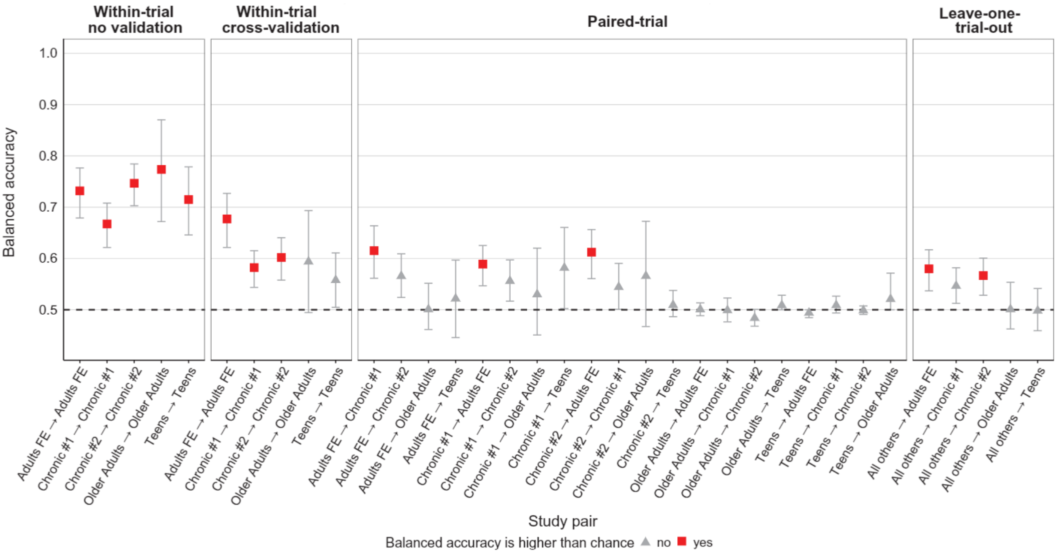
NEUROSCIENCE

Illusory generalizability of clinical prediction models

Adam M. Chekroud[1,2]*, Matt Hawrilenko[1], Hieronimus Loho[2], Julia Bondar[1], Ralitza Gueorguieva[3], Alkomiet Hasan[4], Joseph Kambeitz[5], Philip R. Corlett[2], Nikolaos Koutsouleris[6], Harlan M. Krumholz[7], John H. Krystal[2], Martin Paulus[8]

It is widely hoped that statistical models can improve decision-making related to medical treatments. Because of the cost and scarcity of medical outcomes data, this hope is typically based on investigators observing a model's success in one or two datasets or clinical contexts. We scrutinized this optimism by examining how well a machine learning model performed across several independent clinical trials of antipsychotic medication for schizophrenia. Models predicted patient outcomes with high accuracy within the trial in which the model was developed but performed no better than chance when applied out-of-sample. Pooling data across trials to predict outcomes in the trial left out did not improve predictions. These results suggest that models predicting treatment outcomes in schizophrenia are highly context-dependent and may have limited generalizability.

Real-world example: Chekroud et al., Science 383, 164–167 (2024)



Section 2

Using K-Fold CV on 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

CMSE381_F2025_Schedule : Schedule

	M	9/22	***Project Day & Review***		
	W	9/24	**Midterm #1**		
12	F	9/26	Leave one out CV	5.1.1, 5.1.2	
13	M	9/29	k-fold CV	5.1.3	
14	W	10/1	More k-fold CV	5.1.4-5	
15	F	10/3	k-fold CV for classification	5.1.5	
16	M	10/6	Subset selection	6.1	
17	W	10/8	Shrinkage: Ridge	6.2.1	
18	F	10/10	Shrinkage: Lasso	6.2.2	HW #4 Due Sun 10/12
19	M	10/13	PCA	6.3	
20	W	10/15	PCR	6.3	
	F	10/17	***Review***		
	M	10/20	Fall Break		
	W	10/22	**Midterm #2**		
21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/28
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	W	10/29	Regression Splines	7.4	