Ch 6.3: PCR

Lecture 20 - CMSE 381

Prof. Mengsen Zhang

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

Wed, Oct 15, 2025

Announcements

Last time:

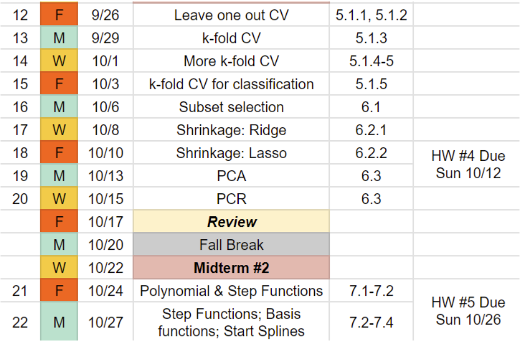
- PCA

This lecture:

- PCR

Announcements:

- Submit your questions by Thursday (10/16) 5 pm!
- Exam #2 on next week!
  - Bring 8.5x11 sheet of paper
  - Handwritten both sides
  - Anything you want on it, but must be your work
  - You will turn it in
  - Non-internet calculator
  - questions about project



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 10/26
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	

Section 1

Previously. . .

Shrinkage

Find β to minimize

$$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

subject to:

Least Squares:

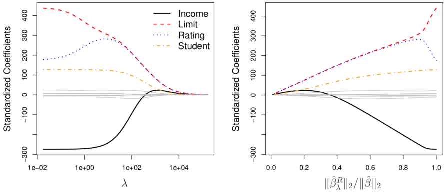
No constraints

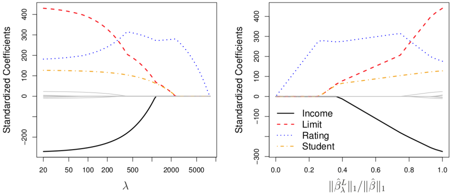
Ridge:

$$\sum_{j=1}^{p} \beta_j^2 \leq s$$

The Lasso:

$$\sum_{j=1}^{p} |\beta_j| \leq s$$





Linear transformation of predictors

Original Predictors:

$$X_1, \cdots, X_p$$

New Predictors:

$$Z_1, \cdots, Z_M$$

$$Z_m = \sum_{j=1}^{p} \varphi_{jm} X_j$$

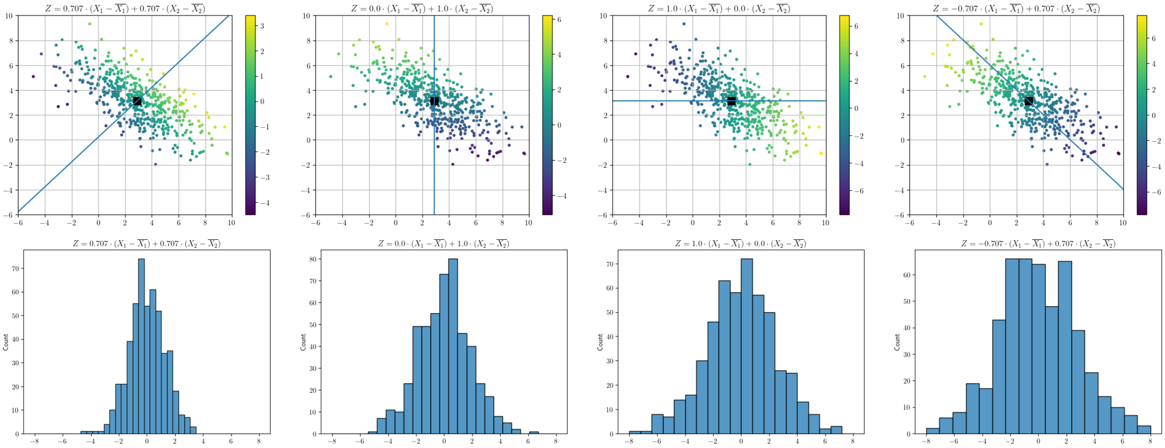
The goal:

- Find good φ's (PCA)
- Fit regression model on Z_i's using least squares (PLS)

$$y_i = \theta_0 + \sum_{m=1}^{M} \theta_m z_{im} + \varepsilon_i$$

- Hope that lower dimensions means less overfitting

PCA - First PC



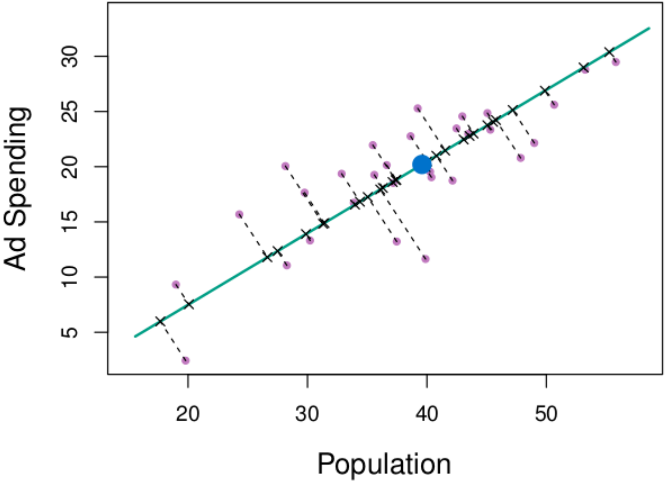
$Z = 0.707 \cdot (X_1 - \overline{X_1}) + 0.707 \cdot (X_2 - \overline{X_2})$

$Z = 0.0 \cdot (X_1 - \overline{X_1}) + 1.0 \cdot (X_2 - \overline{X_2})$

$Z = 1.0 \cdot (X_1 - \overline{X_1}) + 0.0 \cdot (X_2 - \overline{X_2})$

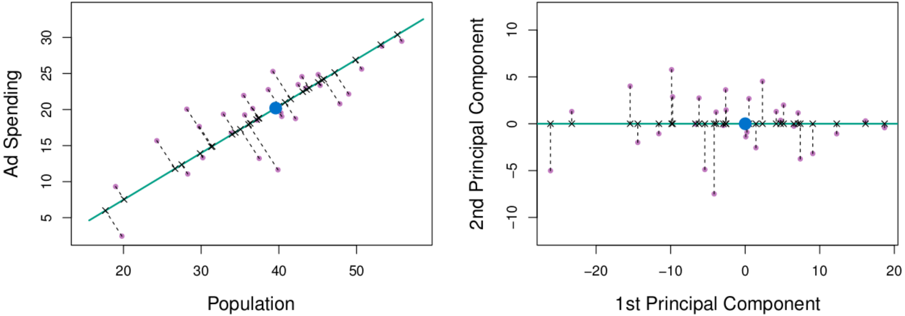
$Z = -0.707 \cdot (X_1 - \overline{X_1}) + 0.707 \cdot (X_2 - \overline{X_2})$

Projection onto first PC



$$Z_1 = 0.839 \cdot (\mathtt{pop} - \overline{\mathtt{pop}}) + 0.544 \cdot (\mathtt{ad} - \overline{\mathtt{ad}})$$

Drawing points in PC space



What will you learn from this lecture?

- Why do you want to use the PCs in regression models?
  - What assumptions do you have to make for it to be a good idea to use **principal component regression (PCR)**?
  - Or conversely, what is a typical bad scenario to use PCR?
- How do you implement PCR in Python?
- How do you interpret the model coefficients when using PCR?
- How do you choose the number of PC to use in PCR?
  - Given a figure of, e.g., cross-validation score as a function of the number of PCs, you should be able to choose appropriately and provide rationales in terms of bias-variance tradeoff.
  - You should be able to generate such figures in Python.
- What is the relationship/differences between PCR and feature selection and regularization methods that you learned in this part of the course?

Section 2

Principal Components Regression

So you've found your PCA coefficients

Now what?

What are we assuming?

Interpretation of PCR coefficients

Original Predictors:

$$X_1, \cdots, X_p$$

New Predictors:

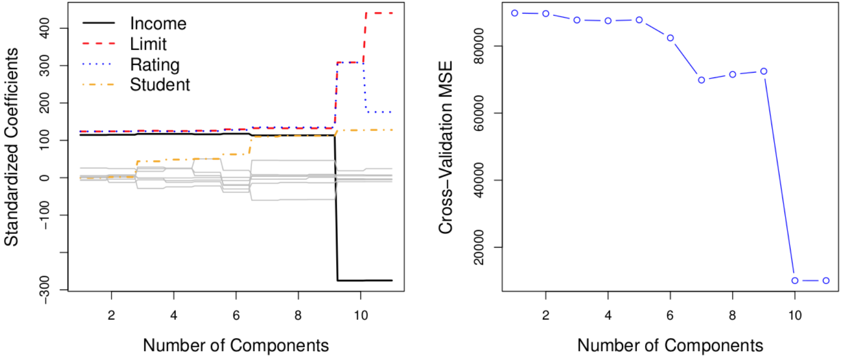
$$Z_1, \cdots, Z_M$$

$$Z_m = \sum_{j=1}^{p} \varphi_{jm} X_j$$

Learned model:

$$y = \theta_0 + \theta_1 Z_1 + \cdots + \theta_M Z_M$$

Picking M

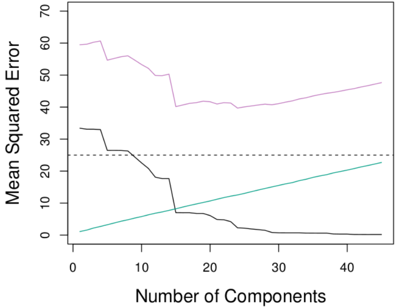

Income
Limit
Rating
Student
Standardized Coefficients
400
300
200
100
0
−100
−300
2
4
6
8
10
Number of Components
Cross−Validation MSE
80000
60000
40000
20000
2
4
6
8
10
Number of Components

Do PCR with hitters data

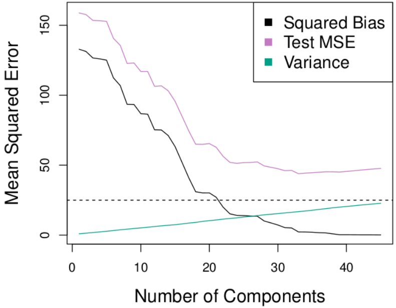
Bias-Variance Trade-off

Example with simulated data: $n = 50$ observations of $p = 45$ predictors

Y is a function of **all predictors**

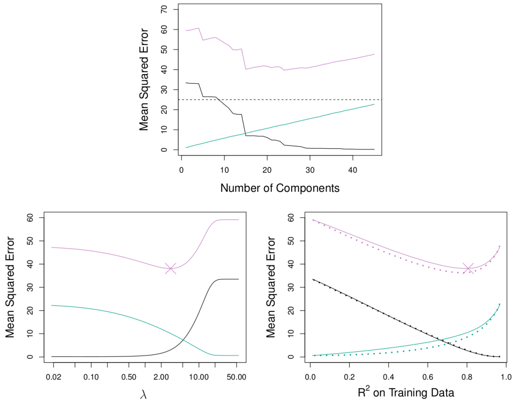


Y is a function of **2 predictors**

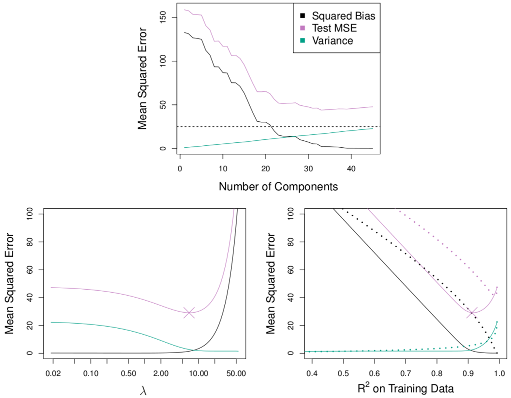


Comparison to results on shrinkage

Y is a function of all predictors



Y is a function of 2 predictors



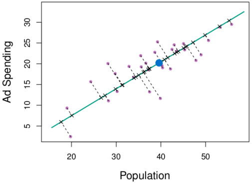
Properties of PCR

Test your understanding: PollEv

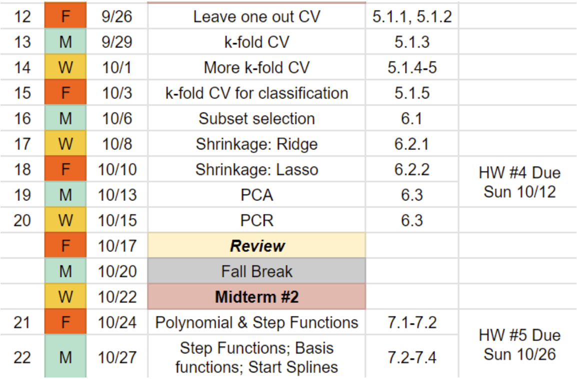
TL;DR

PCR

- Unsupervised dimensionality reduction + linear regression
- Choose component Z_1 in the direction of most variance using only X_i's information
- Choose Z_2 and beyond by the same method after "getting rid" of info in the directions already explained



Next time



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 10/26
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	