

Ch 6.3: PCR

Lecture 20 - CMSE 381

Prof. Guanqun Cao

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Wed, Oct 15, 2025

Last time:

- PCA

This lecture:

- PCR

Announcements:

- Exam #2 on Monday!
 - ▶ 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

Section 1

Previously...

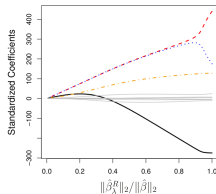
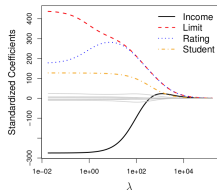
Shrinkage

Find β to minimize

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

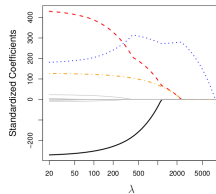
subject to:

Least Squares:
No constraints



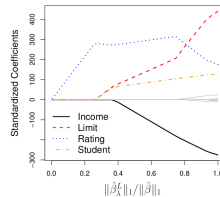
Ridge:

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



The Lasso:

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



Linear transformation of predictors

Original Predictors:

$$X_1, \dots, X_p$$

New Predictors:

$$Z_1, \dots, Z_M$$

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

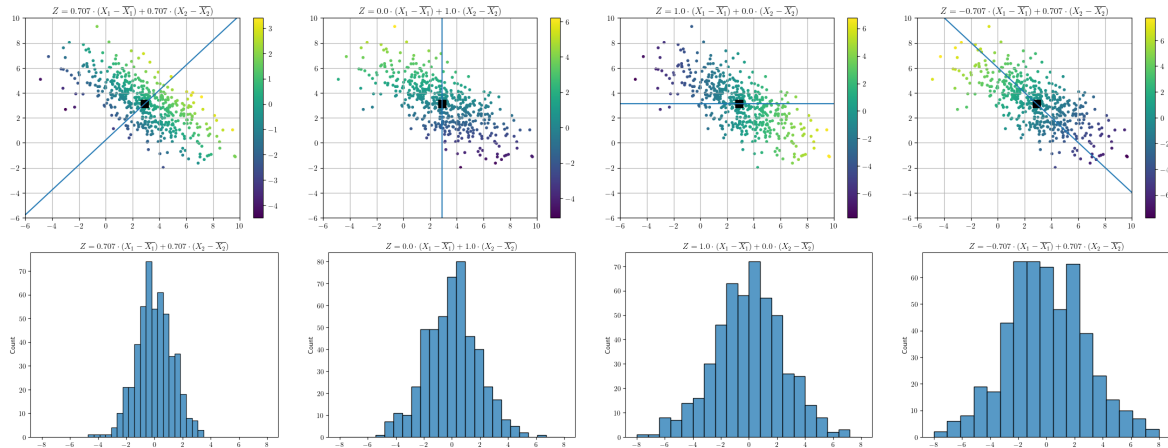
The goal:

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

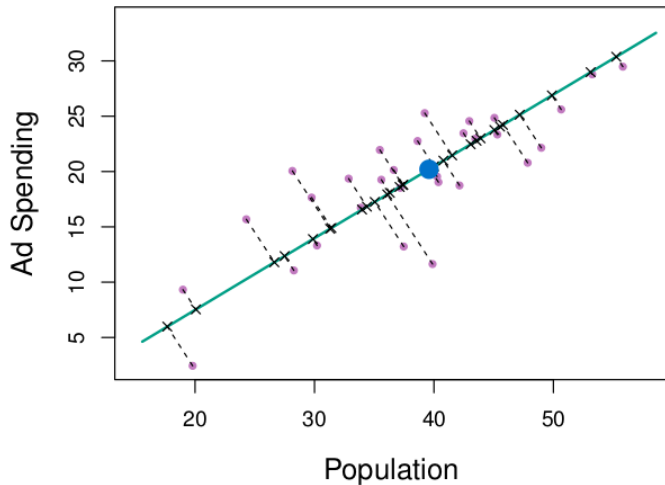
$$y_i = \theta_0 + \sum_{m=1}^M \theta_m z_{im} + \varepsilon_i$$

- Hope that lower dimensions means less overfitting

PCA - First PC

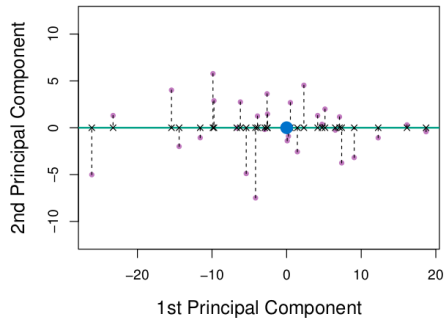
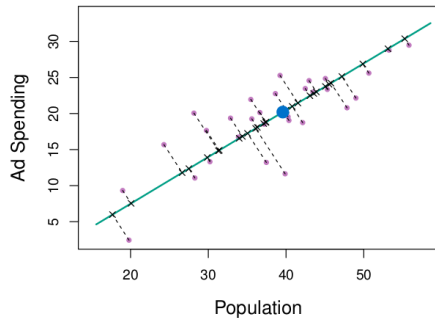


Projection onto first PC



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

Drawing points in PC space



Section 2

Principal Components Regression

So you've found your PCA coefficients

Now what?

What are we assuming?

Interpretation of PCR coefficients

Original Predictors:

$$X_1, \dots, X_p$$

New Predictors:

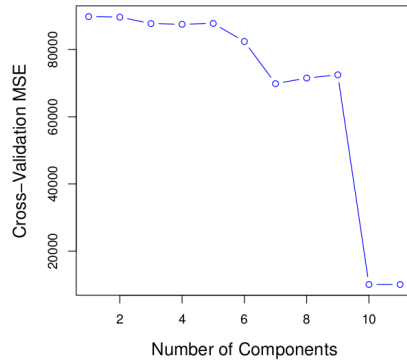
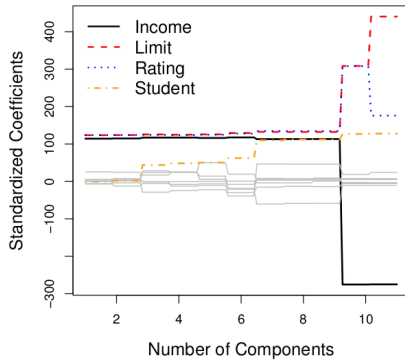
$$Z_1, \dots, Z_M$$

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

Learned model:

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

Picking M

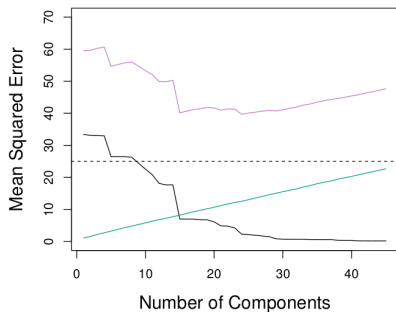


Do PCR with hitters data

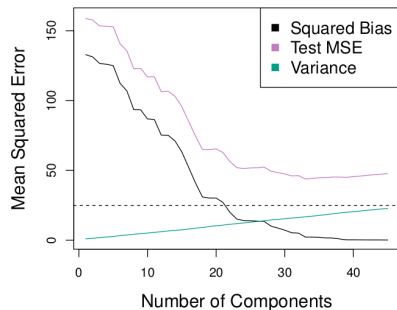
Bias-Variance Trade-off

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

Y is a function of **all** predictors

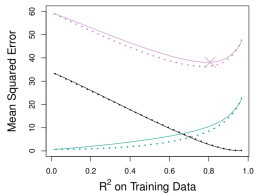
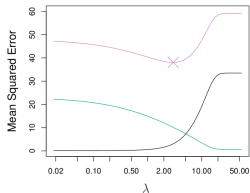
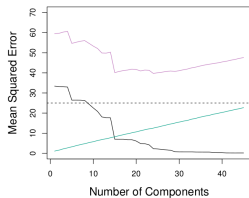


Y is a function of **2** predictors

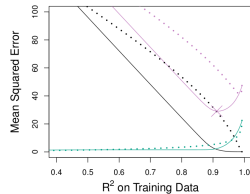
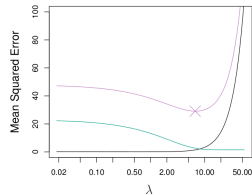
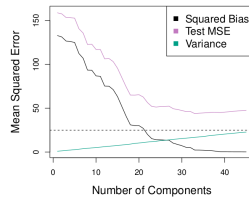


Comparison to results on shrinkage

Y is a function of all predictors



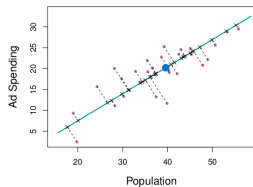
Y is a function of 2 predictors



Properties of PCR

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

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	