Ch 6.3: PCR
Lecture 20 - CMSE 381

Prof. Elizabeth Munch

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

Fri, Oct 18, 2024

Announcements

Last time:

- PCA

This lecture:

- PCR

Announcements:

- Exam #2 on Friday!
  - ▶ 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 if you want it

| Lec #	Date			Reading	HW
12	Mon	9/30	Leave one out CV	5.1.1, 5.1.2	
13	Wed	10/2	k-fold CV	5.1.3	
14	Fri	10/4	More k-fold CV,	5.1.4-5	
15	Mon	10/7	k-fold CV for classification	5.1.5	
16	Wed	10/9	Subset selection	6.1	HW #4 Due Weds 10/9
17	Fri	10/11	Shrinkage: Ridge	6.2.1	
18	Mon	10/14	Shrinkage: Lasso	6.2.2	
19	Wed	10/16	Dimension Reduction	6.3	
20	Fri	10/18	Overflow, Possibly more dimension reduction?		HW #5 Due Fri 10/18
	Mon	10/21	No class - Fall break		
	Wed	10/23	**Review**		
	Fri	10/25	**Midterm #2**		

Section 1

Last time

Shrinkage

Find β to minimize

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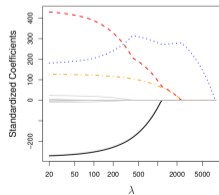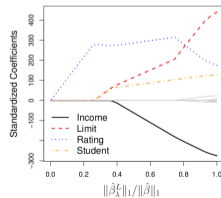
subject to:

Least Squares:
No constraints
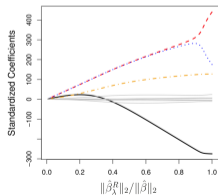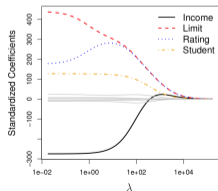
Ridge:

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

The Lasso:

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

Linear transformation of predictors

Original Predictors:

$$X_1, \cdots, X_p$$

New Predictors:
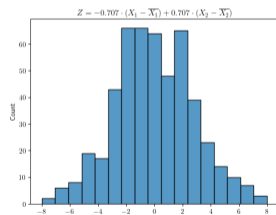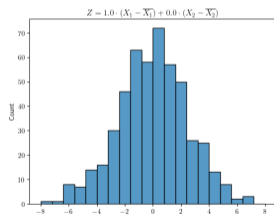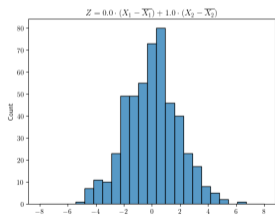
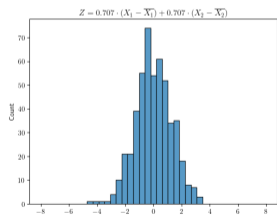$$Z_1, \cdots, Z_M$$
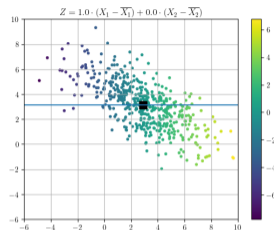
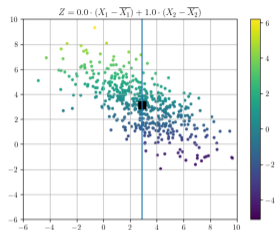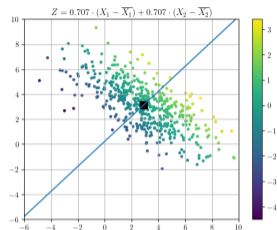$$Z_m = \sum_{j=1}^{p} \varphi_{jm} X_j$$

The goal:

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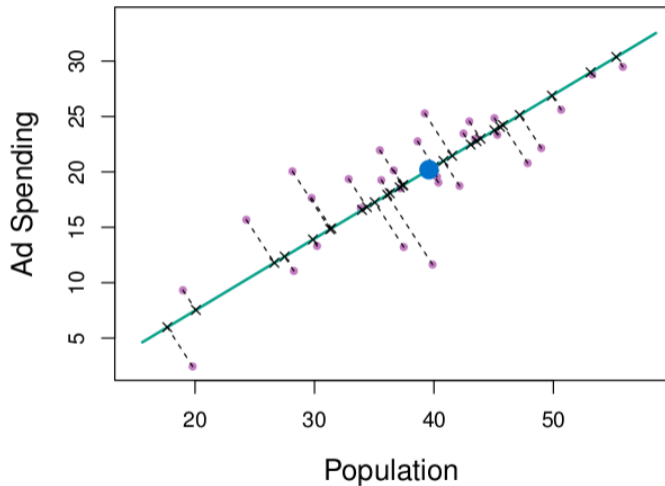- Fit regression model on Z_i's using least squares (PLS)

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

- Hope that lower dimensions means less overfitting
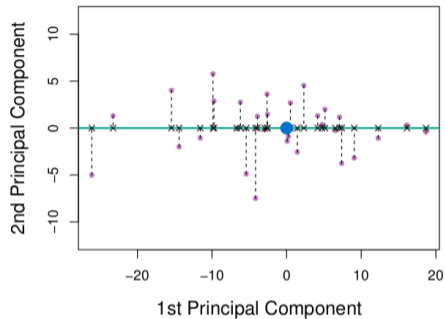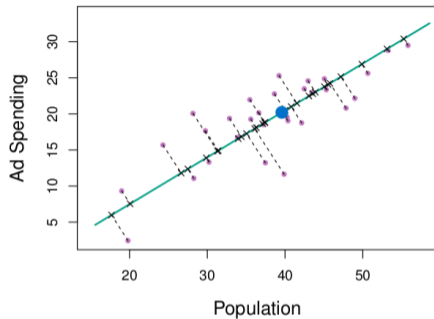
PCA - First PC
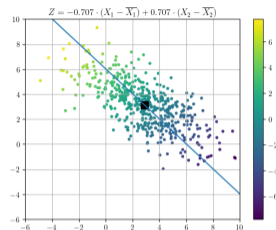
Projection onto first PC

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

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

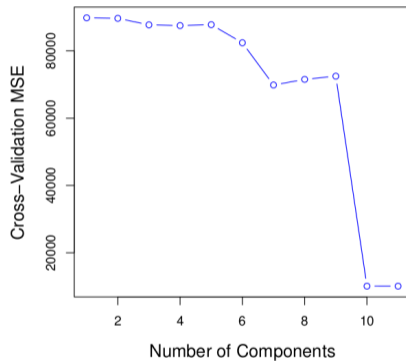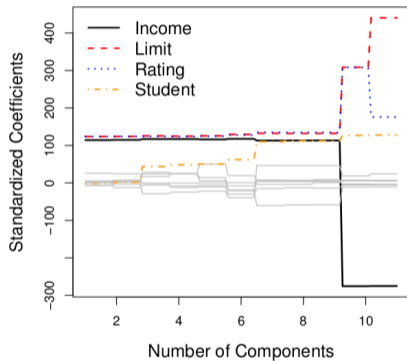**Original Predictors:**

$$X_1, \cdots, X_p$$

New Predictors:

$$Z_1, \cdots, Z_M$$

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

Learned model:

$$y = \theta_0 + \theta_1 Z_1 + \cdots + \theta_M Z_M$$
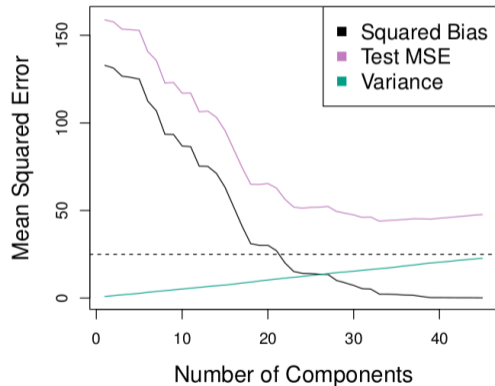
Picking M
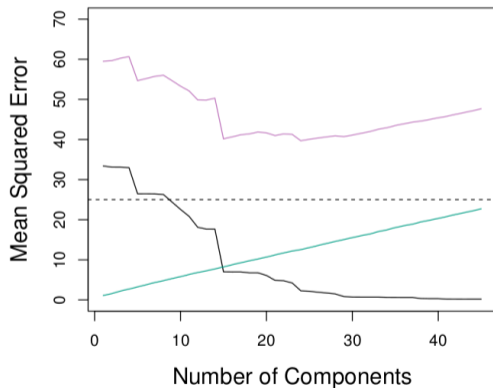
Do PCR with hitters data

Doing better

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

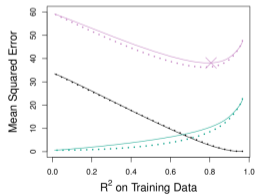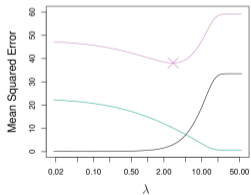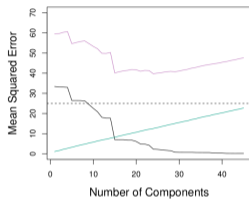Y is a function of **2 predictors**

Doing better

Example with simulated data: $n = 50$ observations of $p = 45$ predictors
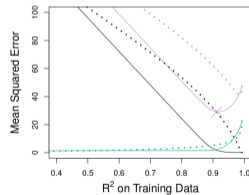Y is a function of **all predictors**

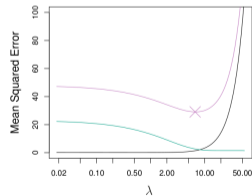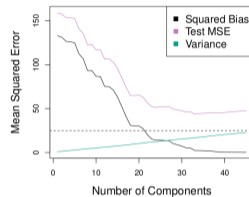# Comparison to results on shrinkage

Y is a function of all predictors

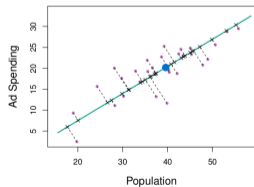Y is a function of 2 predictors

Properties of PCR

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

Lec #	Date				Reading	HW
12	Mon	9/30	Leave one out CV		5.1.1, 5.1.2	
13	Wed	10/2	k-fold CV		5.1.3	
14	Fri	10/4	More k-fold CV,		5.1.4-5	
15	Mon	10/7	k-fold CV for classification		5.1.5	
16	Wed	10/9	Subset selection		6.1	HW #4 Due Weds 10/9
17	Fri	10/11	Shrinkage: Ridge		6.2.1	
18	Mon	10/14	Shrinkage: Lasso		6.2.2	
19	Wed	10/16	Dimension Reduction		6.3	
20	Fri	10/18	Overflow, Possibly more dimension reduction?			HW #5 Due Fri 10/18
	Mon	10/21	No class - Fall break			
	Wed	10/23	**Review**			
	Fri	10/25	**Midterm #2**			