Ch 6.3: Dimension Reduction

Lecture 18 - CMSE 381

Prof. Elizabeth Munch

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

Weds, Oct 26, 2022

Announcements

Last time:

- Shrinkage: Ridge and Lasso

This lecture:

- PCA / PCR
- PLS

Announcements:

- Homework #6 due Friday

Section 1

Last time

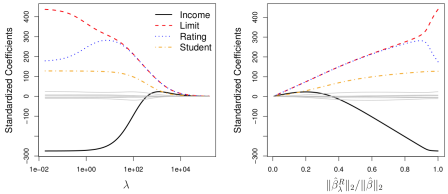
Shrinkage

Find β to minimize

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

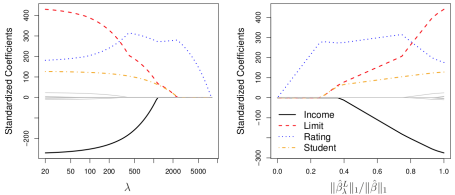
subject to:

Least Squares:
No constraints



Ridge:

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



The Lasso:

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

Section 2

Dimension Reduction

Linear transformation of predictors

Original Predictors:

$$X_1, \cdots, X_p$$

New Predictors:

$$Z_1, \cdots, X_M$$

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

An example or two

The goal

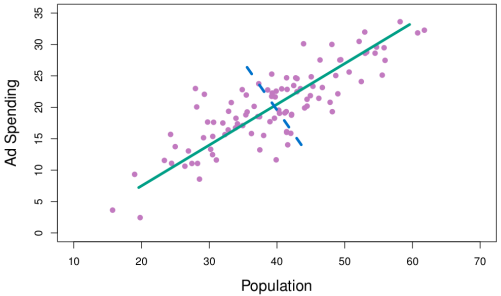
- Find good φ's
- Fit regression model on Z_i's using least squares

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

Section 3

PCA

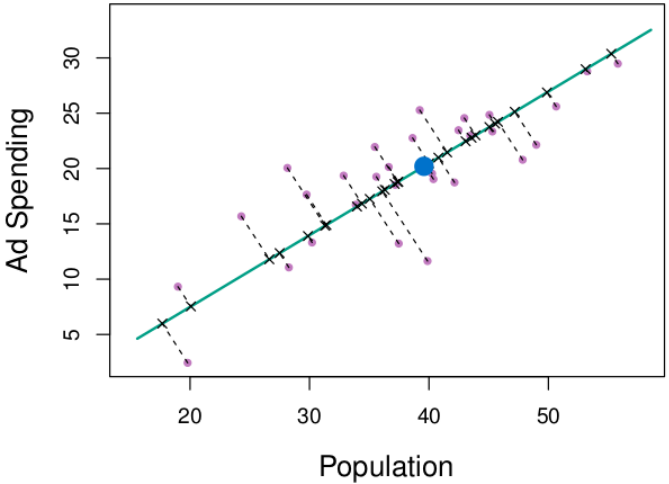
An example dataset



Projection onto a line

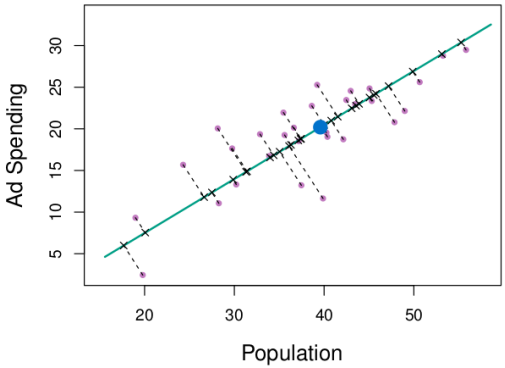
https://www.desmos.com/calculator/cih7wy8oyg

Projection onto first PC



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

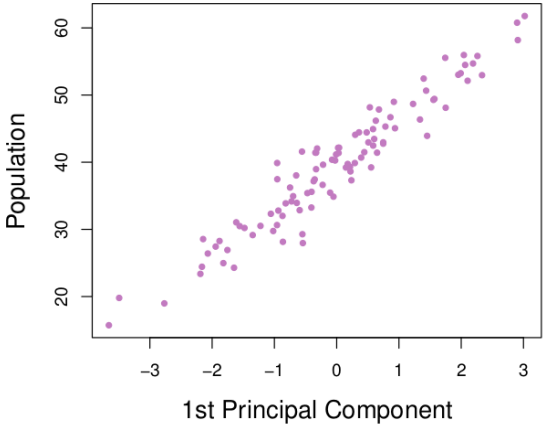
What does it mean to have the highest variance

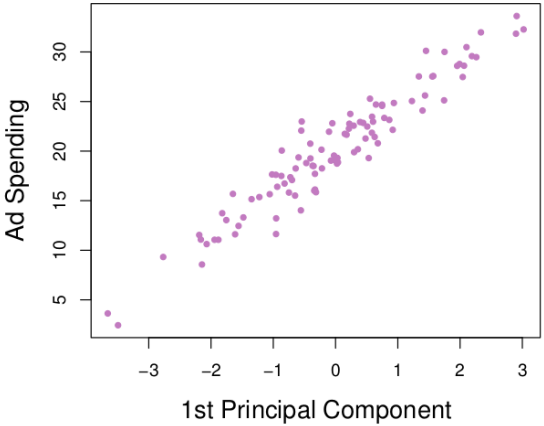


Toy for learning PCA

https://www.desmos.com/calculator/qq14tyjz0z

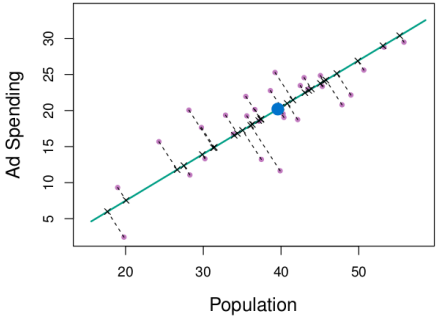
Principal component scores

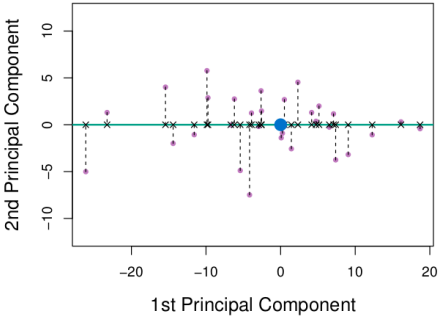




$$z_{i1} = 0.839 \cdot (\text{pop}_i - \overline{\text{pop}}) + 0.544 \cdot (\text{ad}_i - \overline{\text{ad}})$$

Another view





The other principal components

Do PCA with Penguins

Section 4

Principal Components Regression

So you've found your PCA coefficients

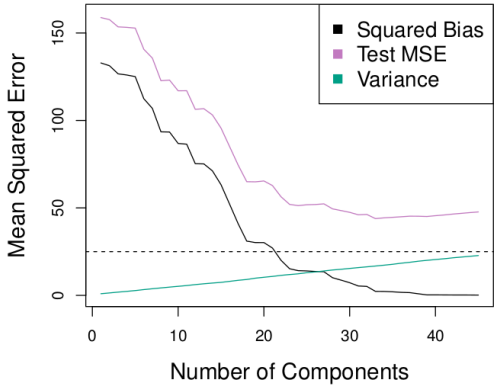
Now what?

What are we assuming?

Doing better

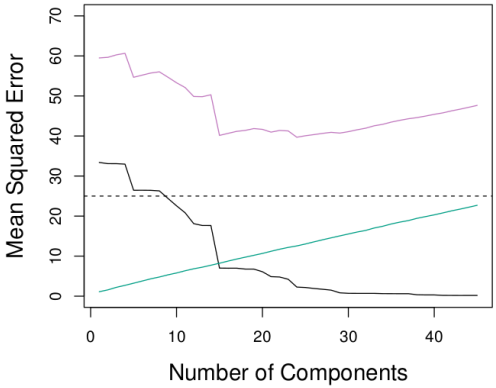
Example with simulated data: $n = 50$ observations of $p = 45$ predictors

Y is a function of 2 predictors



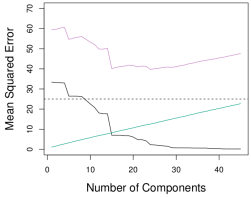
Doing better

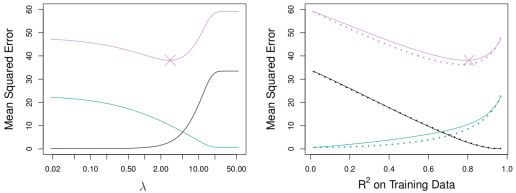
Example with simulated data: $n = 50$ observations of $p = 45$ predictors
Y is a function of all predictors



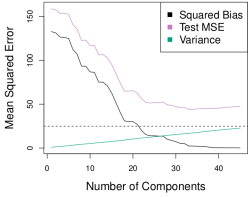
Comparison to results on shrinkage

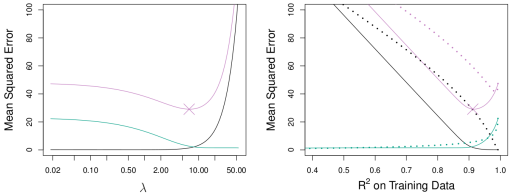
Y is a function of all predictors



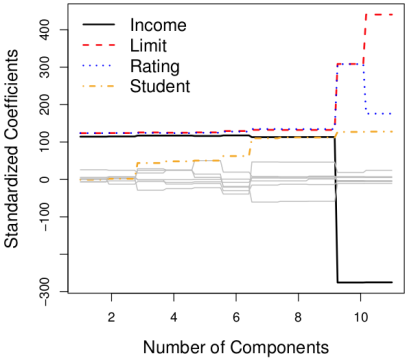


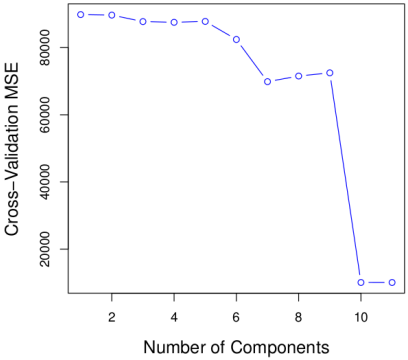
Y is a function of 2 predictors





Picking M


Income
Limit
Rating
Student
Standardized Coefficients
−300
−100
0
100
200
300
400
2
4
6
8
10
Number of Components


Cross−Validation MSE
20000
40000
60000
80000
2
4
6
8
10
Number of Components

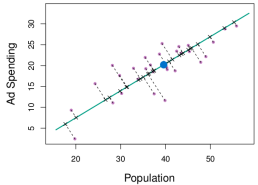
Properties of PCR

Do PCR with hitters data

TL;DR

PCA

- Unsupervised dimensionality reduction
- 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



PCR

- Do PCA on input data
- Do Linear Regression on chosen number of PCs.
- Warning: Lose interpretability of the coefficients.

Next time

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