Ch 6.3: Dimension Reduction

Lecture 20 - CMSE 381

Prof. Elizabeth Munch

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

Weds, Oct 27, 2023

Announcements

Last time:

- Exam #2

This lecture:

- PCA / PCR
- PLS

Announcements:

- Homework #6 ..... posted?

Section 1

Last time

Shrinkage

Find β to minimize:

Least Squares:

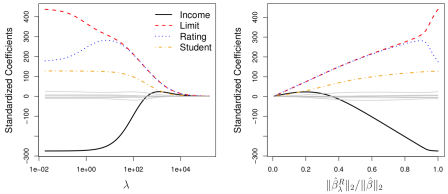
$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

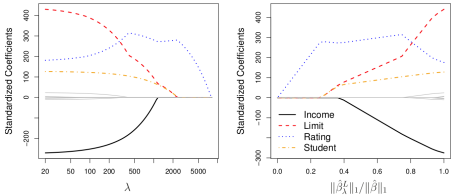
Ridge:

$$RSS + \sum_{j=1}^{p}\beta_j^2$$

The Lasso:

$$RSS + \sum_{j=1}^{p}|\beta_j|$$





Section 2

Dimension Reduction

Linear transformation of predictors

Original Predictors:

$$X_1, \cdots, X_p$$

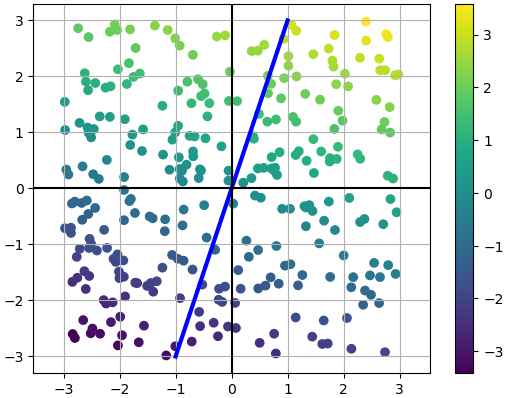
New Predictors:

$$Z_1, \cdots, X_M$$

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

An example or two

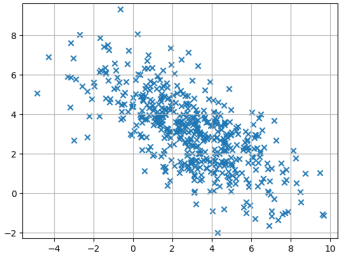
Geometric interpretation

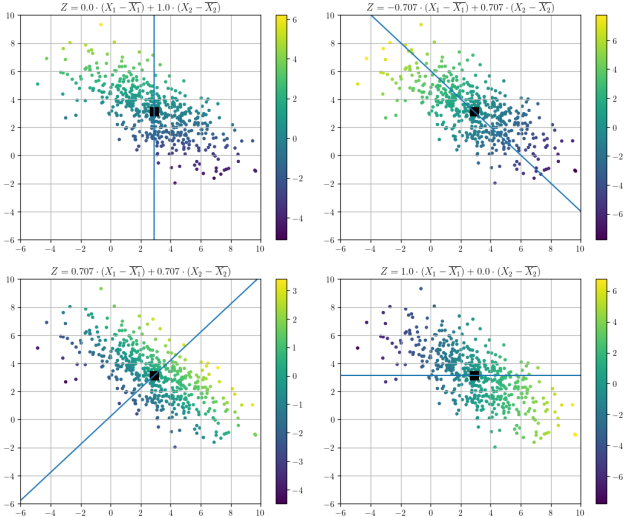


Projection onto a line

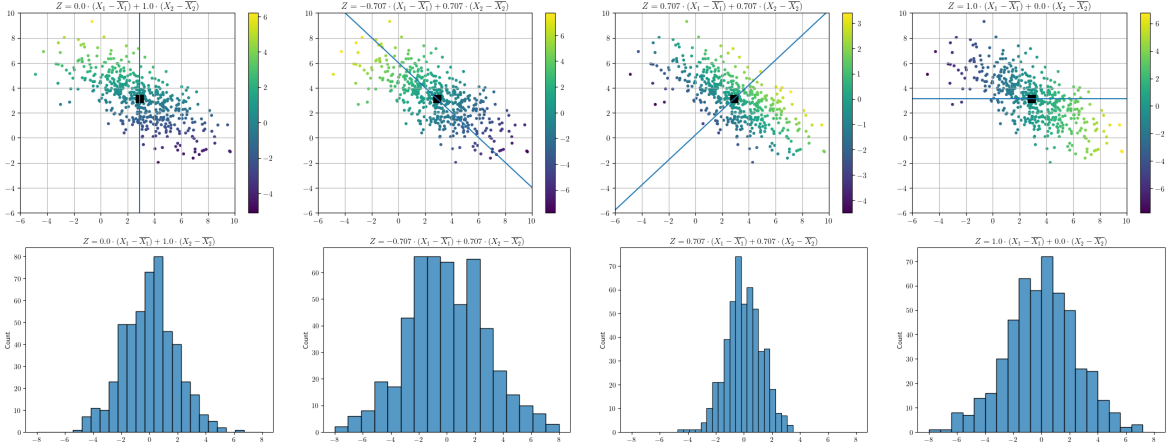
https://www.desmos.com/calculator/cih7wy8oyg

Different projections





Histograms of Z values



The goal

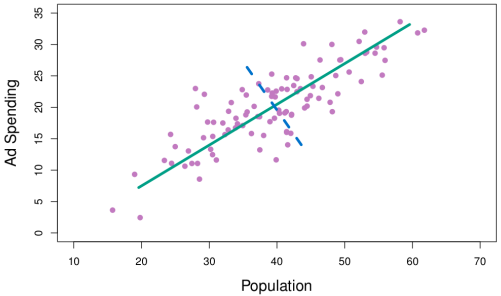
- Find good φ's for some $M \ll p$
- Fit regression model on Z_i's using least squares

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

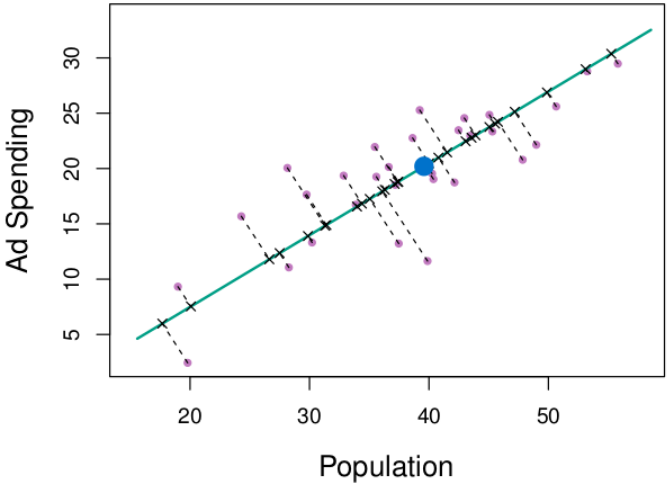
Section 3

PCA

An example dataset

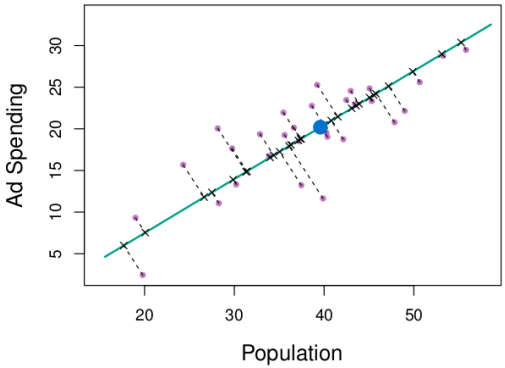


Projection onto first PC



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

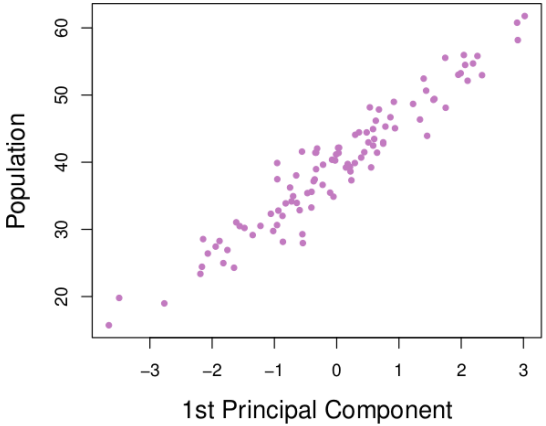
What does it mean to have the highest variance

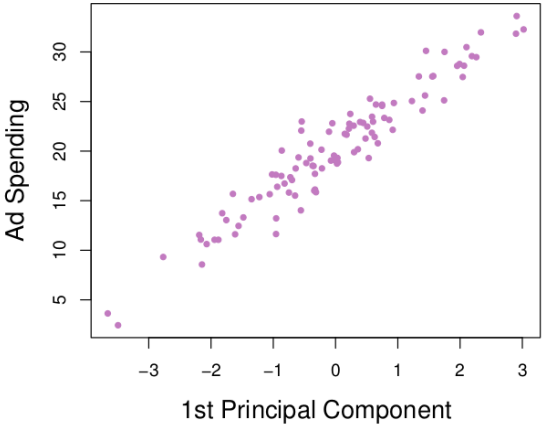


Toy for learning PCA

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

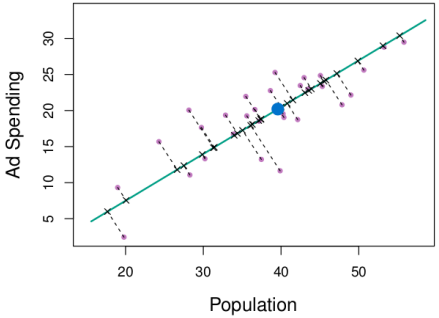
Principal component scores

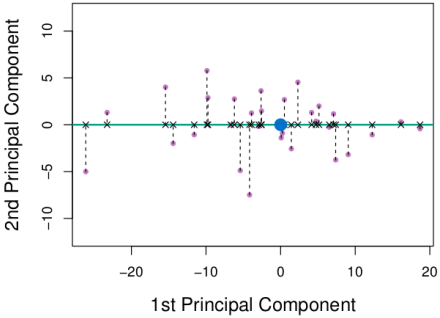




$$z_{i1} = 0.839 \cdot (\text{pop}_i - \overline{\text{pop}}) + 0.544 \cdot (\text{ad}_i - \overline{\text{ad}})$$

Another view





The other principal components

Do PCA with Penguins

Section 4

Principal Components Regression

So you've found your PCA coefficients

Now what?

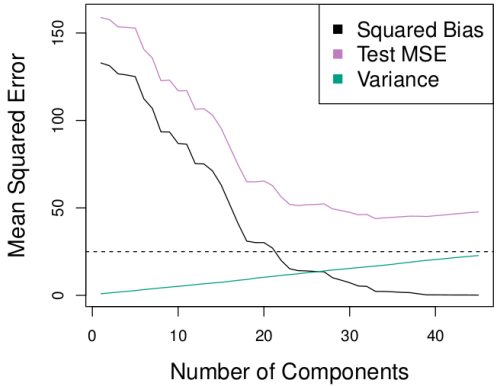
What are we assuming?

Do PCR with hitters data

Doing better

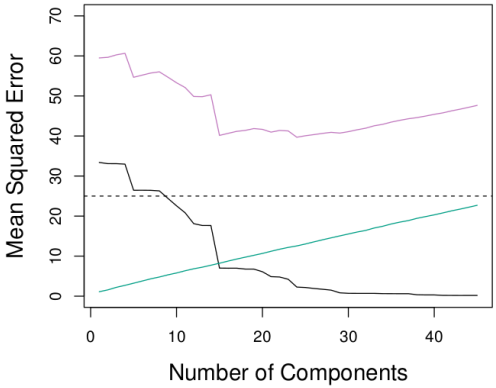
Example with simulated data: $n = 50$ observations of $p = 45$ predictors

Y is a function of 2 predictors



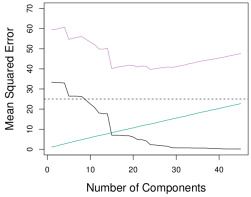
Doing better

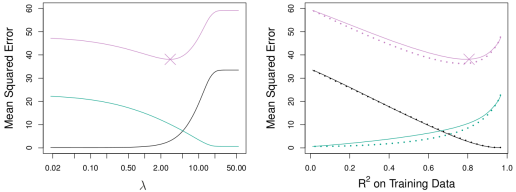
Example with simulated data: $n = 50$ observations of $p = 45$ predictors
Y is a function of all predictors



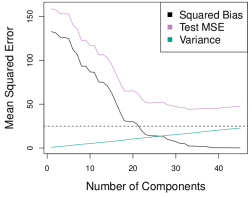
Comparison to results on shrinkage

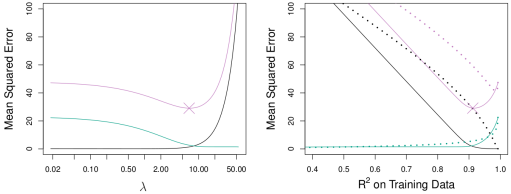
Y is a function of all predictors



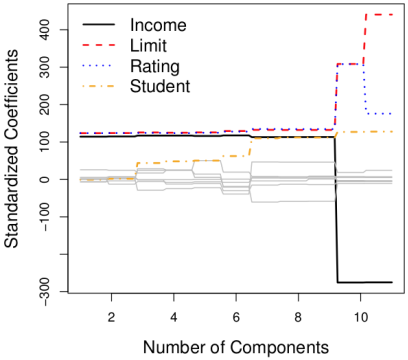


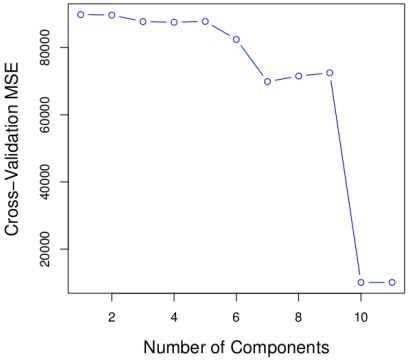
Y is a function of 2 predictors





Picking M


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

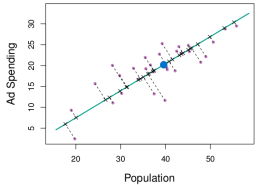

Cross−Validation MSE
20000
40000
60000
80000
2
4
6
8
10
Number of Components

Properties of PCR

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

Lec #	Date			Reading	Homeworks
20	Fri	Oct 27	Dimension Reduction	6.3	
21	Mon	Oct 30	More dimension reduction; High dimensions	6.4	
22	Wed	Nov 1	Polynomial & Step Functions	7.1,7.2	
23	Fri	Nov 3	Step Functions	7.2	HW #6 Due
24	Mon	Nov 6	Basis functions, Regression Splines	7.3,7.4	
25	Wed	Nov 8	Decision Trees	8.1	
26	Fri	Nov 10	Random Forests	8.2.1, 8.2.2	
27	Mon	Nov 13	Maximal Margin Classifier	9.1	
28	Wed	Nov 15	SVC	9.2	
29	Fri	Nov 17	SVM	9.3, 9.4	
30	Mon	Nov 20	Single layer NN	10.1	
31	Wed	Nov 22	Overflow/project day?		
	Fri	Nov 24	No class - Thanksgiving		
	Mon	Nov 27	**Review**		
	Wed	Nov 29	**Midterm #3**		
32	Fri	Dec 1	Multi Layer NN	10.2	
33	Mon	Dec 4	CNN	10.3	
34	Wed	Dec 6	Unsupervised Learning & Clustering	12.1, 12.4	
35	Fri	Dec 8	Overflow/Project day?		Project due