Ch 6.3: Dimension Reduction - PCA

Lecture 19 - CMSE 381

Prof. Mengsen Zhang

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

Mon, March 10, 2025

Announcements

Last time:

- Shrinkage: Ridge and Lasso

This lecture:

- PCA

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
- Project: by Exam # 2
  - project partner
  - ideas about what method to use

	W	2/12	**Midterm #1**		
12	F	2/14	Leave one out CV	5.1.1, 5.1.2	
13	M	2/17	k-fold CV	5.1.3	
14	W	2/19	More k-fold CV	5.1.4-5	
15	F	2/21	k-fold CV for classification	5.1.5	
16	M	2/24	Subset selection	6.1	
17	W	2/26	Shrinkage: Ridge	6.2.1	
18	F	2/28	Shrinkage: Lasso	6.2.2	HW #4 Due Sun 3/2
	M	3/3	Spring Break		
	W	3/5	Spring Break		
	F	3/7	Spring Break		
19	M	3/10	PCA	6.3	
20	W	3/12	PCR	6.3	
	F	3/14	***Review***		HW #5 Due Sun 3/16
	M	3/17	**Midterm #2**		

Section 1

Last time

Goal

- Fit model using all p predictors
- Aim to constrain (regularize) coefficient estimates
- Shrink the coefficient estimates towards 0

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Shrinkage

Find β to minimize:

Least Squares:

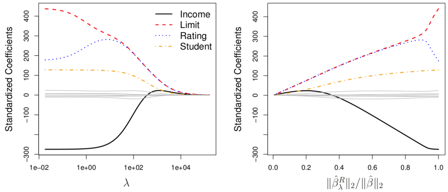
$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

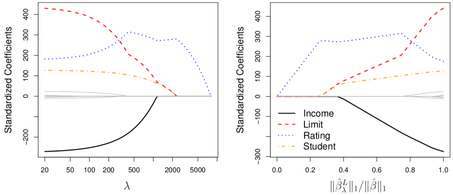
Ridge:

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

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:

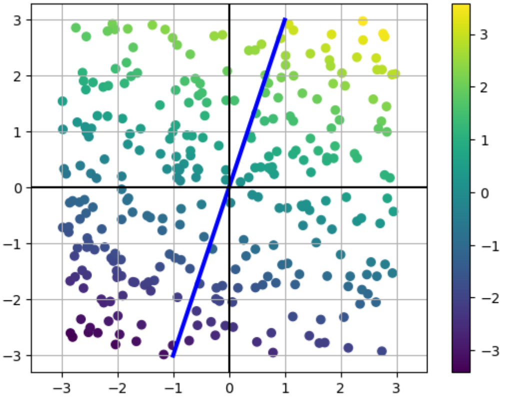
$$Z_1, \cdots, Z_M$$

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

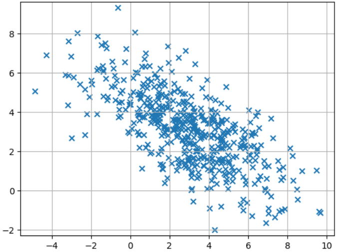
An example or two

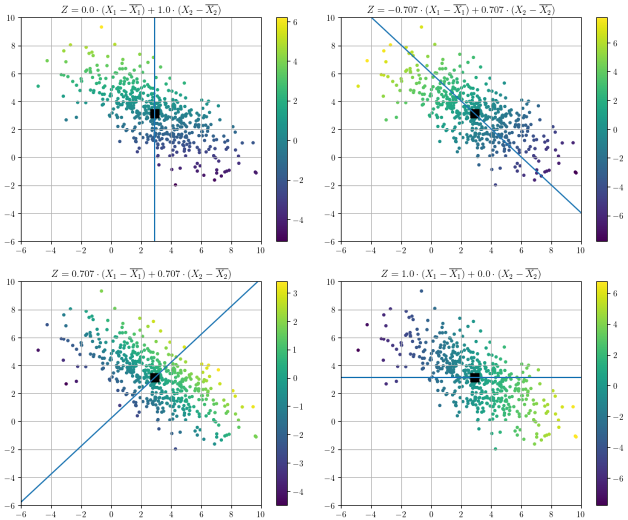
Geometric interpretation

- projection on a line

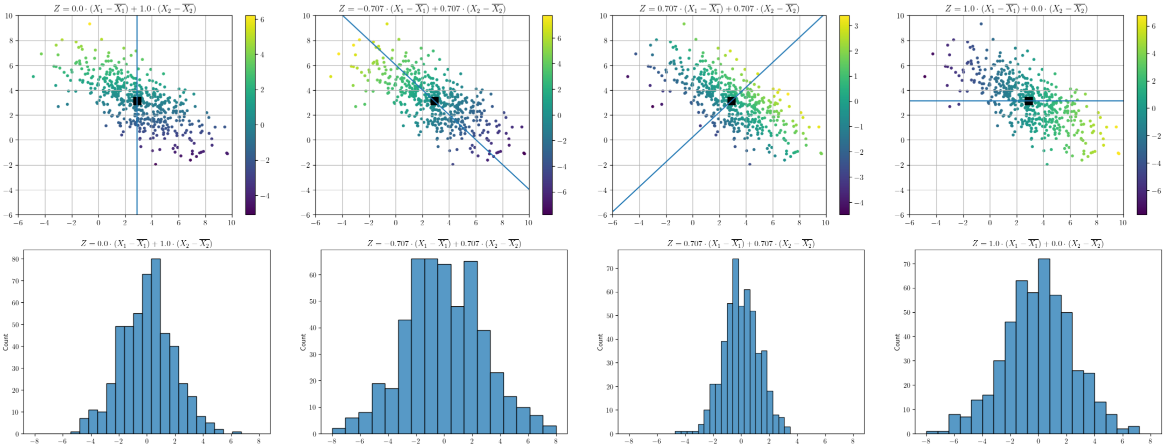


Different projections





Histograms of Z values



The goal

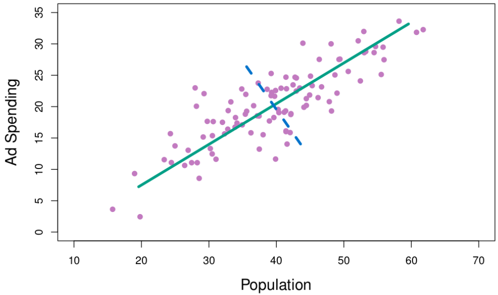
- Find good φ's for some $M \ll p$
- Fit regression model on Z_i's using least squares

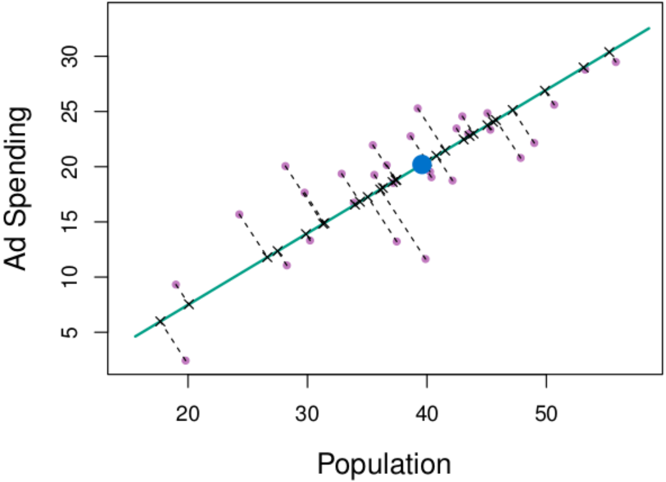
$$y_i = \theta_0 + \sum_{m=1}^{M} \theta_m z_{im} + \varepsilon_i$$

Section 3

PCA

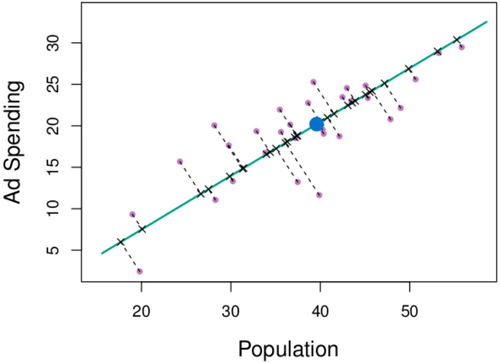
An example dataset





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

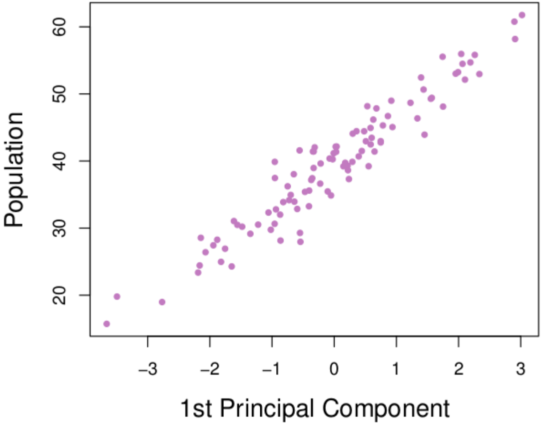
What does it mean to have the highest variance

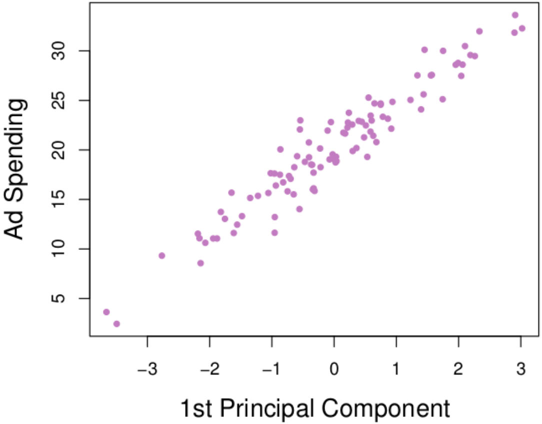


Toy for learning PCA

https://www.desmos.com/calculator/gvmq07pg1k

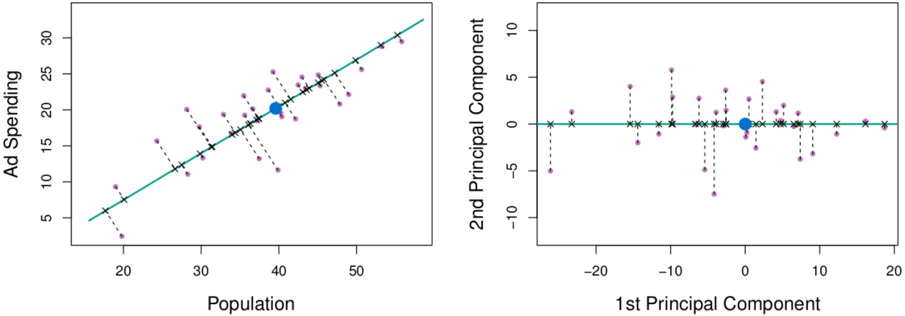
Principal component scores





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

Another view

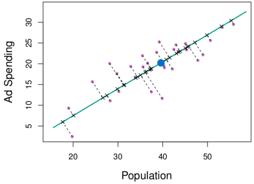


The other principal components

Do PCA with Penguins

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



Next time

	W	2/12	**Midterm #1**		
12	F	2/14	Leave one out CV	5.1.1, 5.1.2	
13	M	2/17	k-fold CV	5.1.3	
14	W	2/19	More k-fold CV	5.1.4-5	
15	F	2/21	k-fold CV for classification	5.1.5	
16	M	2/24	Subset selection	6.1	
17	W	2/26	Shrinkage: Ridge	6.2.1	
18	F	2/28	Shrinkage: Lasso	6.2.2	HW #4 Due Sun 3/2
	M	3/3	Spring Break		
	W	3/5	Spring Break		
	F	3/7	Spring Break		
19	M	3/10	PCA	6.3	
20	W	3/12	PCR	6.3	
	F	3/14	***Review***		HW #5 Due Sun 3/16
	M	3/17	**Midterm #2**		