Ch 6.3: Dimension Reduction - PCA

Lecture 19 - CMSE 381

Prof. Elizabeth Munch

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

Weds, Oct 16, 2024

Announcements

Last time:

- Shrinkage: Ridge and Lasso

This lecture:

- PCA

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

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:

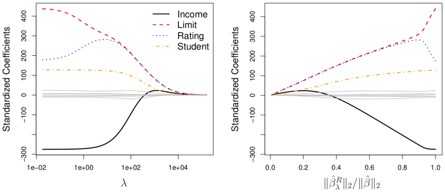
$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

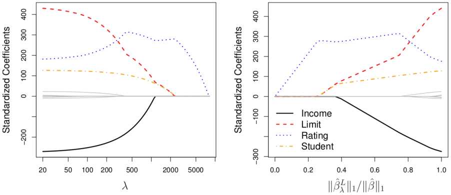
Ridge:

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

The Lasso:

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





Section 2

Dimension Reduction

Linear transformation of predictors

Original Predictors:

$$X_1, \cdots, X_p$$

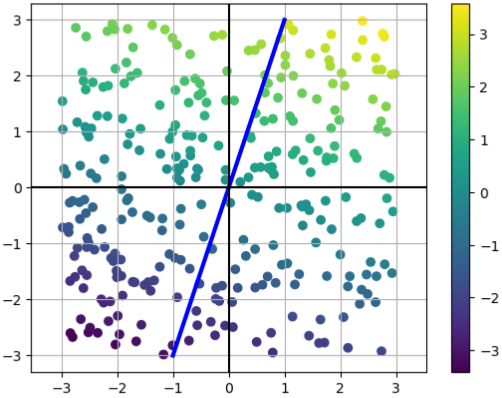
New Predictors:

$$Z_1, \cdots, Z_M$$

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

An example or two

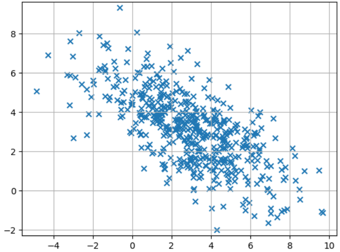
Geometric interpretation

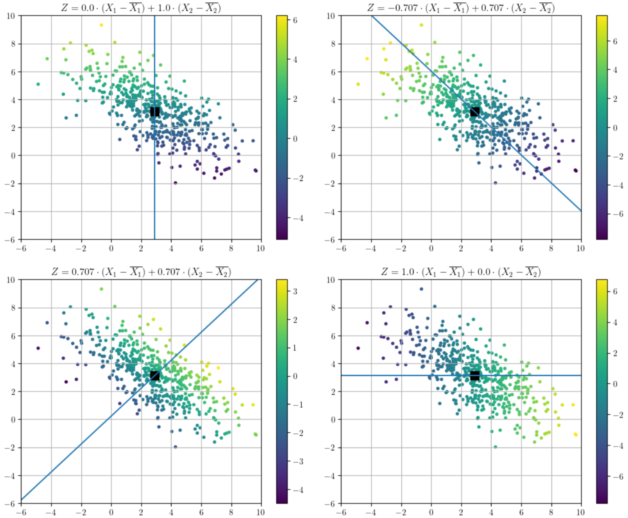


Projection onto a line

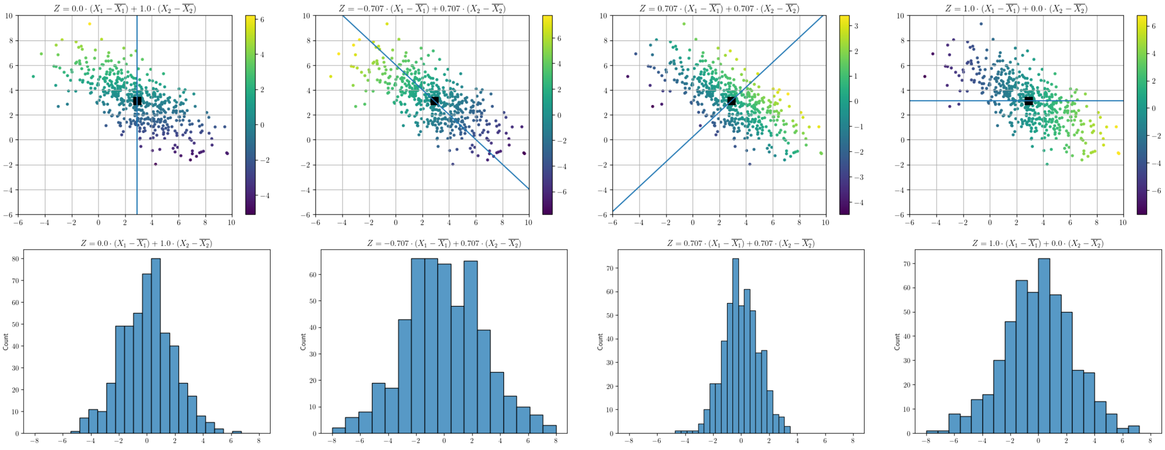
https://www.desmos.com/calculator/cih7wy8oyg

Different projections





Histograms of Z values



The goal

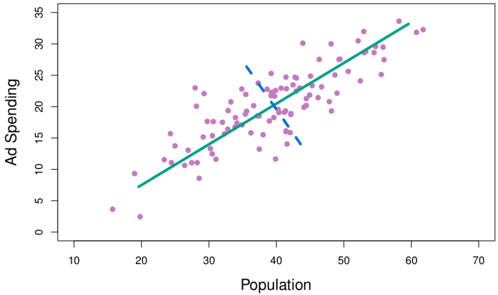
- Find good φ's for some $M \ll p$
- Fit regression model on Z_i's using least squares

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

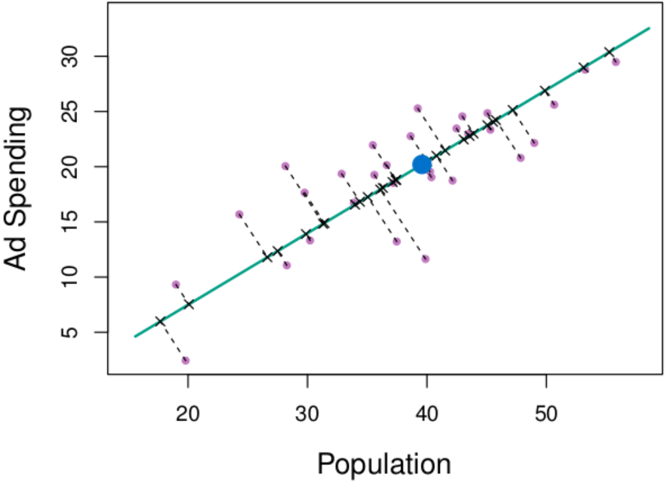
Section 3

PCA

An example dataset

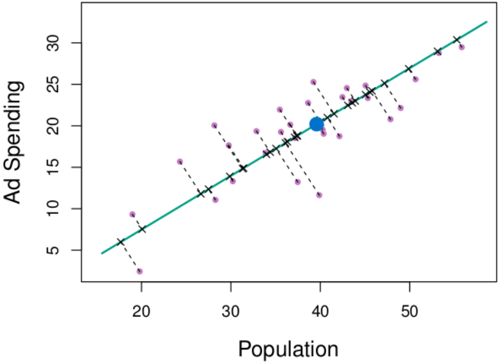


Projection onto first PC



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

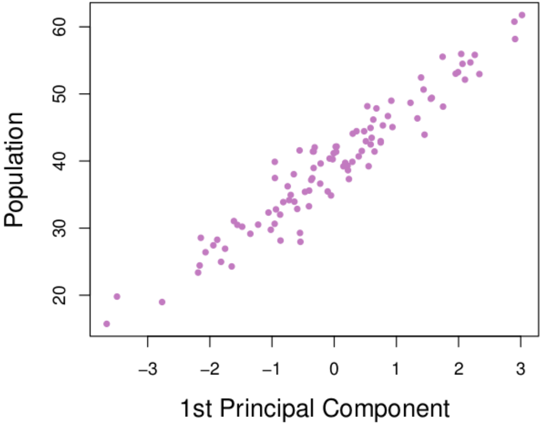
What does it mean to have the highest variance

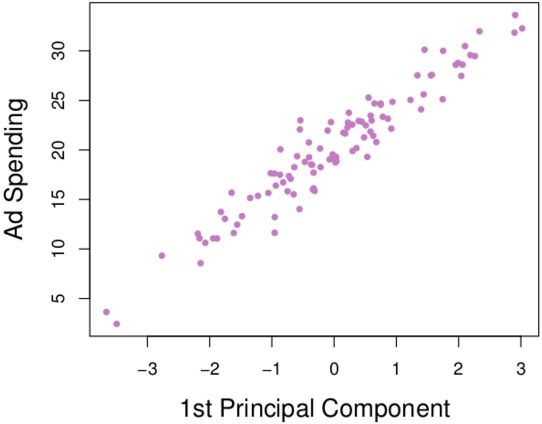


Toy for learning PCA

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

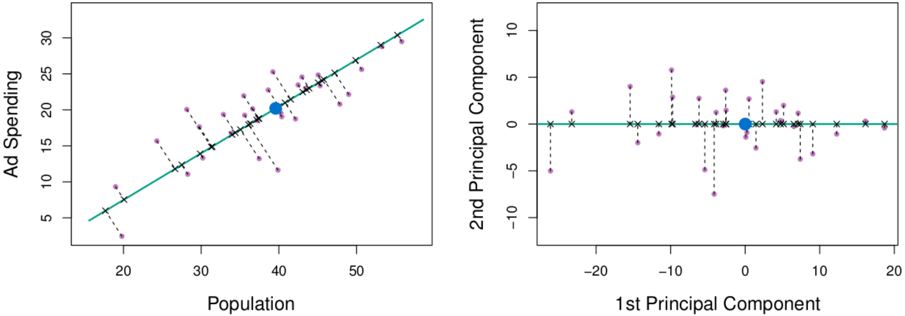
Principal component scores





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

Another view

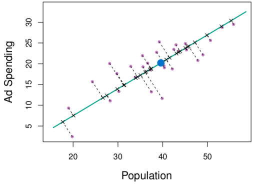


The other principal components

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

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**		