Ch 6.3: Dimension Reduction - PCA

Lecture 19 - CMSE 381

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

Mon, March 9, 2026

Announcements

Last time:

- Shrinkage: Ridge and Lasso

This lecture:

- PCA

Announcements:

- Exam #2 next week (Monday 3/16)!
  - Bring 8.5x11 sheet of paper
  - Handwritten both sides
  - Anything you want on it, but must be your work
  - Write your name and your group number
  - You will turn it in
  - Non-internet calculator
- Project: by Exam # 2
  - project partner, ideas about what method to use

11	F	2/6	Multiple Logistic Regression / Multinomial Logistic Regression	4.3.4-5	HW #2 Due Mon 2/9	
	M	2/9	***Project Day & Review***			
	W	2/11	**Midterm #1**			
12	F	2/13	**Class not held**			
13	M	2/16	Leave one out and k-fold CV	5.1.1-3		
14	W	2/18	More k-fold CV	5.1.4-5		Q5
15	F	2/20	k-fold CV for classification	5.1.5		
16	M	2/23	Subset selection	6.1		
17	W	2/25	Shrinkage: Ridge	6.2.1		
18	F	2/27	Shrinkage: Lasso	6.2.2		
	M	3/2	Spring Break			
	W	3/4	Spring Break			
	F	3/6	Spring Break		HW #3 Due Sun 3/8	
19	M	3/9	PCA	6.3		
20	W	3/11	PCR	6.3		Q6

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:

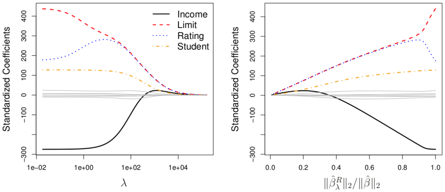
$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

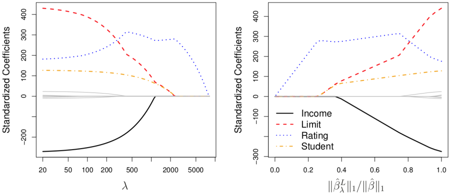
Ridge:

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

The Lasso:

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





What will you learn from this lecture?

- How to create new variables as linear combinations of the original predictors?
- Why do we need Principal Component Analysis (PCA)? What is the main purpose of using it?
- What is a principal component (PC)?
- What does the first PC maximize? You should be able to explain this both geometrically in a plot and mathematically.
- Similarly, what do the subsequent PCs maximize?
- How do you compute the PCs in Python, given a dataset?
- How do you project data points on each PC? You should also be able to plot the data points in the PC space.
- How do you find out how much variance each PC explains?

Section 2

Dimension Reduction

Linear transformation of predictors

Original Predictors:

$$X_1, \cdots, X_p$$

New Predictors:

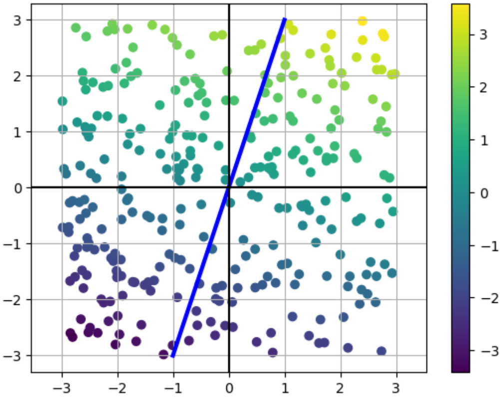
$$Z_1, \cdots, Z_M$$

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

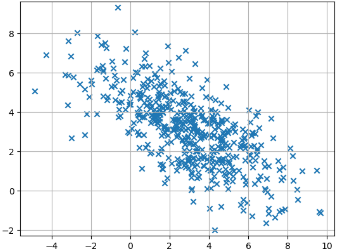
An example or two

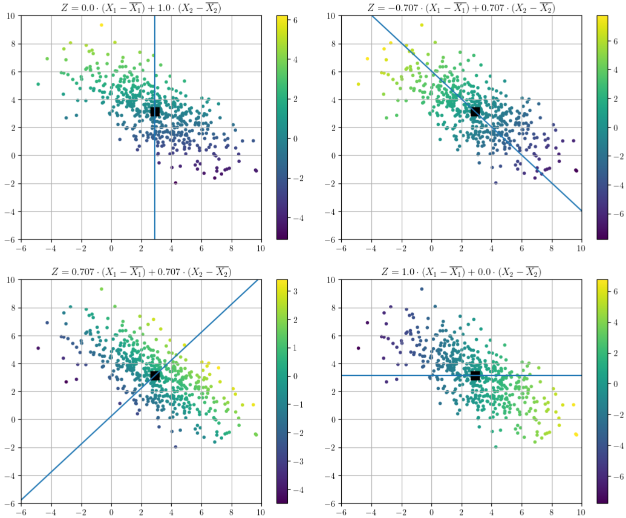
Geometric interpretation

- projection on a line

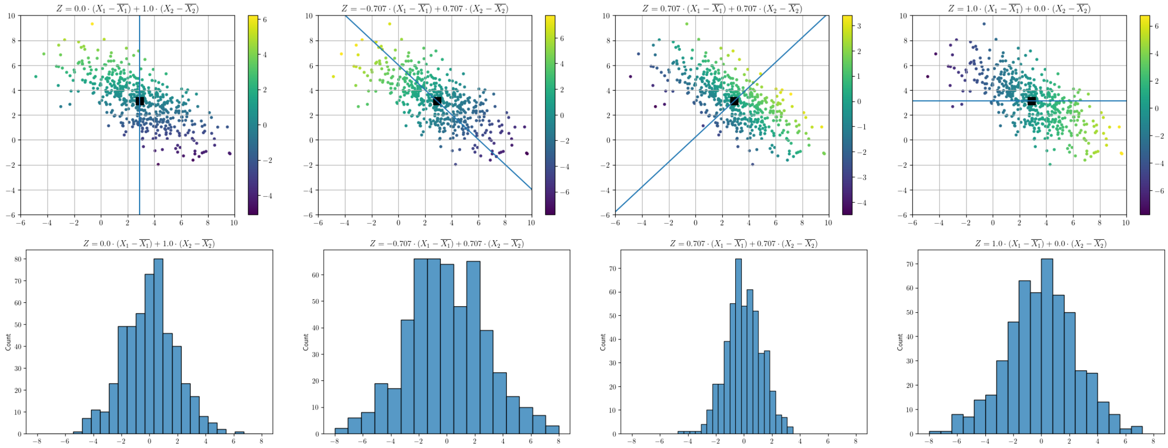


Different projections





Histograms of Z values



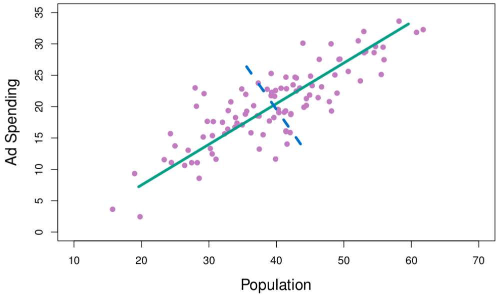
The goal

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

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

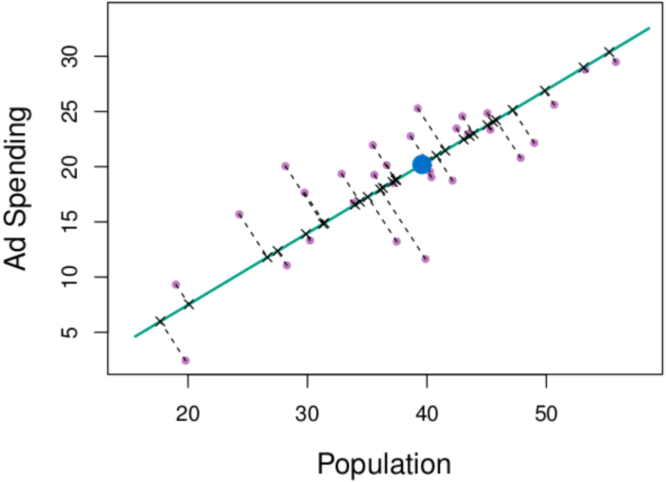
Section 3

PCA



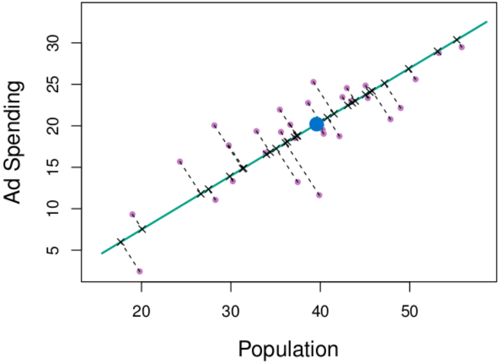
Ad Spending
0
5
10
15
20
25
30
35
10
20
30
40
50
60
70
Population

Projection onto first PC



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

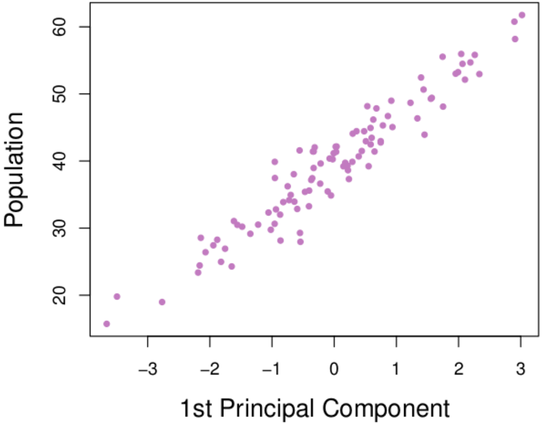
What does it mean to have the highest variance

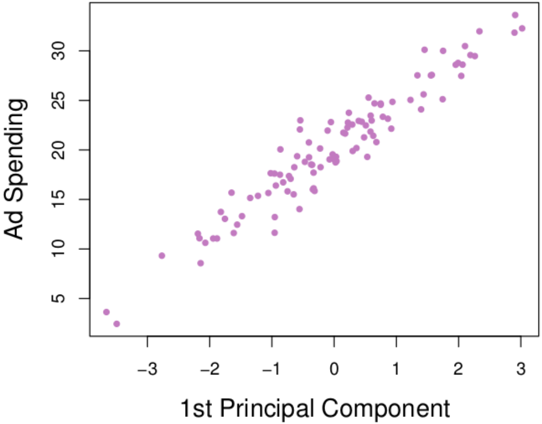


Toy for learning PCA

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

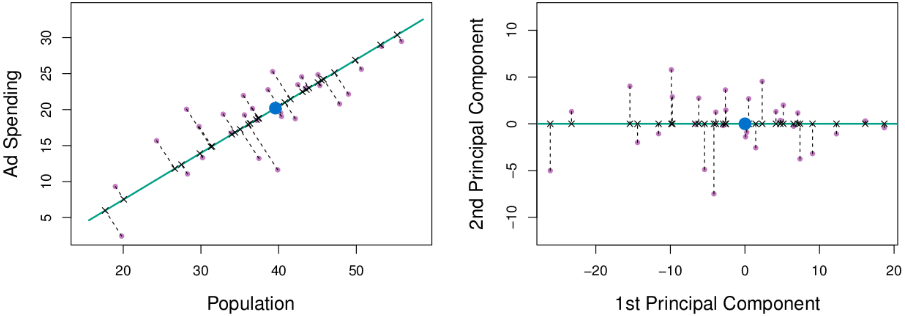
Principal component scores





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

Another view

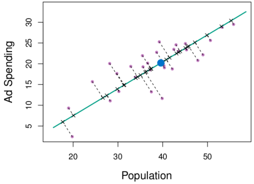


The other principal components

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

11	F	2/6	Multiple Logistic Regression / Multinomial Logistic Regression	4.3.4-5	HW #2 Due Mon 2/9	
	M	2/9	***Project Day & Review***			
	W	2/11	**Midterm #1**			
12	F	2/13	**Class not held**			
13	M	2/16	Leave one out and k-fold CV	5.1.1-3		
14	W	2/18	More k-fold CV	5.1.4-5		Q5
15	F	2/20	k-fold CV for classification	5.1.5		
16	M	2/23	Subset selection	6.1		
17	W	2/25	Shrinkage: Ridge	6.2.1		
18	F	2/27	Shrinkage: Lasso	6.2.2		
	M	3/2	Spring Break			
	W	3/4	Spring Break			
	F	3/6	Spring Break		HW #3 Due Sun 3/8	
19	M	3/9	PCA	6.3		
20	W	3/11	PCR	6.3		Q6