

# Ch 6.3: Dimension Reduction - PCA

## Lecture 19 - CMSE 381

Prof. Guanqun Cao

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Mon, Oct 13, 2025

# Announcements

## Last time:

- Shrinkage: Ridge and Lasso

## This lecture:

- PCA

## Announcements:

- Exam #2 on Wednesday!
  - 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

CMSE381\_F2025\_Schedule : Schedule

|    |   |       |                                                |              |                        |
|----|---|-------|------------------------------------------------|--------------|------------------------|
|    | M | 9/22  | <b>Project Day &amp; Review</b>                |              |                        |
|    | W | 9/24  | <b>Midterm #1</b>                              |              |                        |
| 12 | F | 9/26  | Leave one out CV                               | 5.1.1, 5.1.2 |                        |
| 13 | M | 9/29  | k-fold CV                                      | 5.1.3        |                        |
| 14 | W | 10/1  | More k-fold CV                                 | 5.1.4-5      |                        |
| 15 | F | 10/3  | k-fold CV for classification                   | 5.1.5        |                        |
| 16 | M | 10/6  | Subset selection                               | 6.1          |                        |
| 17 | W | 10/8  | Shrinkage: Ridge                               | 6.2.1        |                        |
| 18 | F | 10/10 | Shrinkage: Lasso                               | 6.2.2        | HW #4 Due<br>Sun 10/12 |
| 19 | M | 10/13 | PCA                                            | 6.3          |                        |
| 20 | W | 10/15 | PCR                                            | 6.3          |                        |
|    | F | 10/17 | <b>Review</b>                                  |              |                        |
|    | M | 10/20 | Fall Break                                     |              |                        |
|    | W | 10/22 | <b>Midterm #2</b>                              |              |                        |
| 21 | F | 10/24 | Polynomial & Step Functions                    | 7.1-7.2      | HW #5 Due<br>Sun 10/28 |
| 22 | M | 10/27 | Step Functions; Basis functions; Start Splines | 7.2-7.4      |                        |
| 23 | W | 10/29 | Regression Splines                             | 7.4          |                        |

# Section 1

Last time

# Goal

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

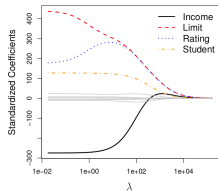
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

# Shrinkage

Find  $\beta$  to minimize:

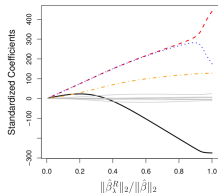
**Least Squares:**

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



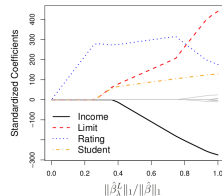
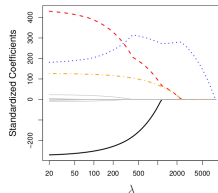
**Ridge:**

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



**The Lasso:**

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



## Section 2

### Dimension Reduction

# Linear transformation of predictors

Original Predictors:

$$X_1, \dots, X_p$$

New Predictors:

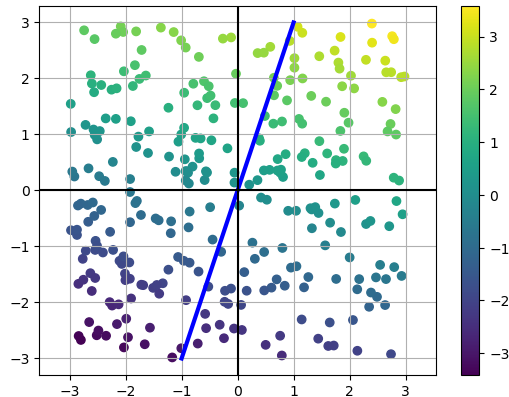
$$Z_1, \dots, Z_M$$

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

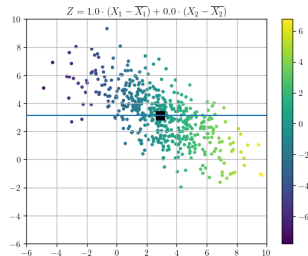
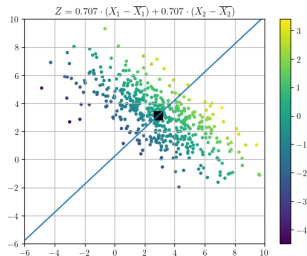
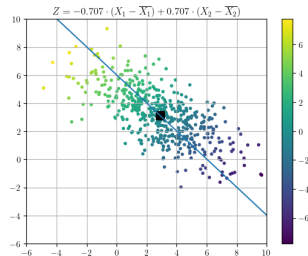
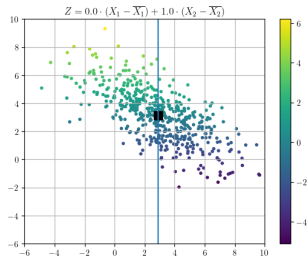
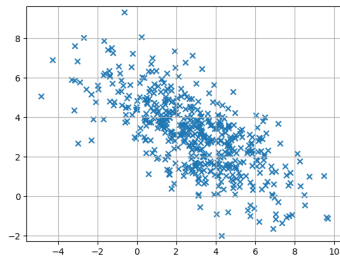
# An example or two

# Geometric interpretation

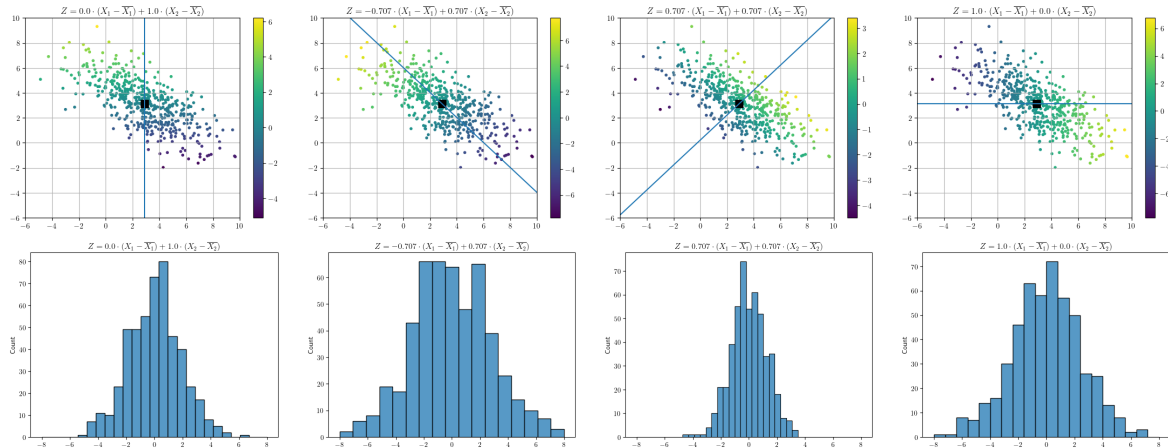
- projection on a line



# Different projections



# Histograms of $Z$ values



# The goal

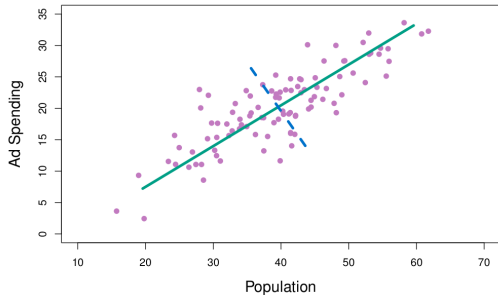
- Find good  $\varphi$ 's for some  $M \ll p$
- Fit regression model on  $Z_i$ 's using least squares

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

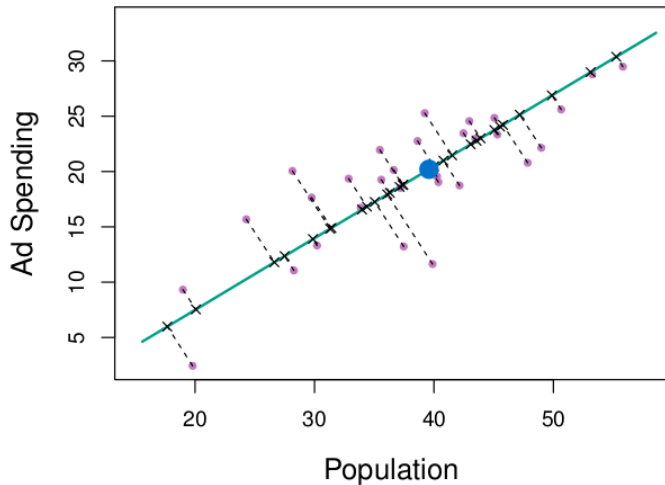
## Section 3

### PCA

# An example dataset

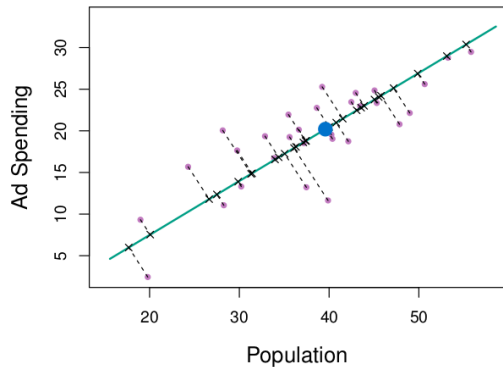


## Projection onto first PC



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

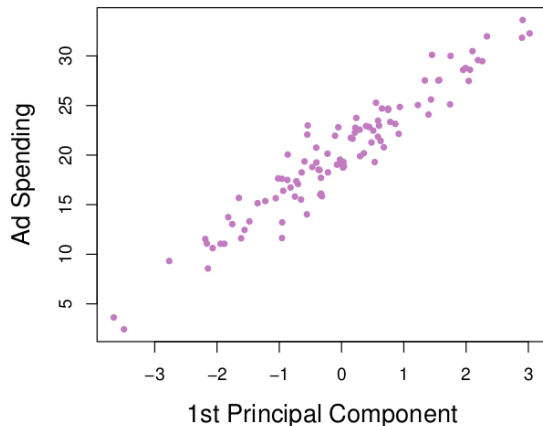
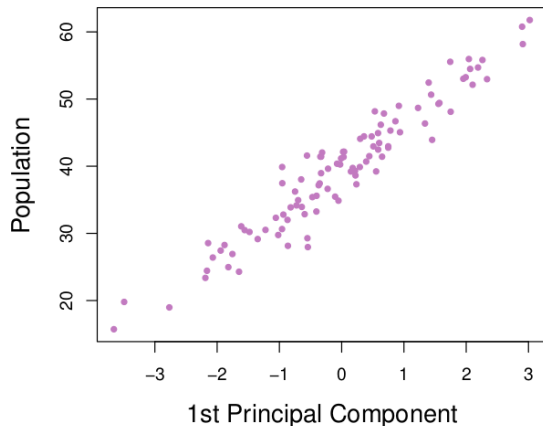
# What does it mean to have the highest variance



# Toy for learning PCA

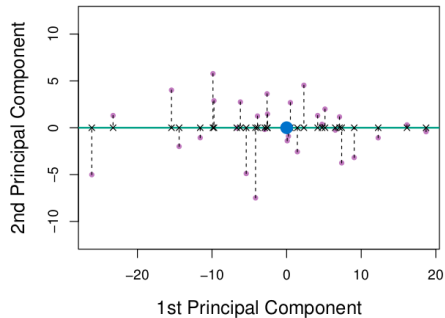
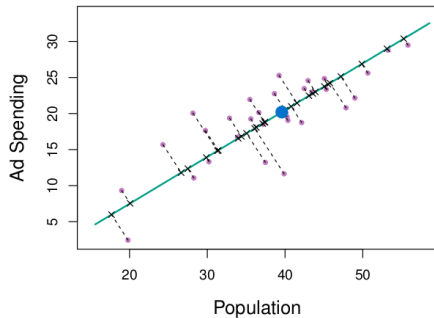
[https://www.desmos.com/  
calculator/gvmq07pg1k](https://www.desmos.com/calculator/gvmq07pg1k)

# 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

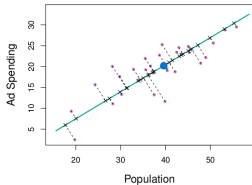


# The other principal components

# Do PCA with Penguins

## 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

## CMSE381\_F2025\_Schedule : Schedule

|    |   |       |                                                |              |                        |
|----|---|-------|------------------------------------------------|--------------|------------------------|
|    | M | 9/22  | <b>Project Day &amp; Review</b>                |              |                        |
|    | W | 9/24  | <b>Midterm #1</b>                              |              |                        |
| 12 | F | 9/26  | Leave one out CV                               | 5.1.1, 5.1.2 |                        |
| 13 | M | 9/29  | k-fold CV                                      | 5.1.3        |                        |
| 14 | W | 10/1  | More k-fold CV                                 | 5.1.4-5      |                        |
| 15 | F | 10/3  | k-fold CV for classification                   | 5.1.5        |                        |
| 16 | M | 10/6  | Subset selection                               | 6.1          |                        |
| 17 | W | 10/8  | Shrinkage: Ridge                               | 6.2.1        |                        |
| 18 | F | 10/10 | Shrinkage: Lasso                               | 6.2.2        | HW #4 Due<br>Sun 10/12 |
| 19 | M | 10/13 | PCA                                            | 6.3          |                        |
| 20 | W | 10/15 | PCR                                            | 6.3          |                        |
|    | F | 10/17 | <b>Review</b>                                  |              |                        |
|    | M | 10/20 | Fall Break                                     |              |                        |
|    | W | 10/22 | <b>Midterm #2</b>                              |              |                        |
| 21 | F | 10/24 | Polynomial & Step Functions                    | 7.1-7.2      | HW #5 Due<br>Sun 10/28 |
| 22 | M | 10/27 | Step Functions; Basis functions; Start Splines | 7.2-7.4      |                        |
| 23 | W | 10/29 | Regression Splines                             | 7.4          |                        |