Ch 6.4: Curse of Dimensionality

Lecture 19 - CMSE 381

Prof. Elizabeth Munch

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

Fri, Oct 28, 2022

Announcements

Last time:

- PCA/PCR

This lecture:

- 6.3: PLS
- 6.4: Issues with higher dimensions

Announcements:

- Homework due Friday
- Monday is review day
- Wednesday is Exam
  - 8.5" x 11" cheat sheet
  - Basic calculator
  - Covers content since last exam (Chapters 5 and 6)

Section 1

Last time

Find β to minimize

$$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

subject to:

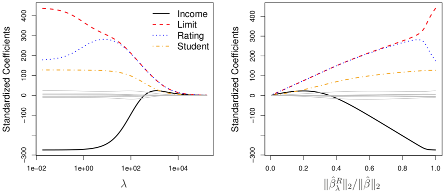
Least Squares:
No constraints

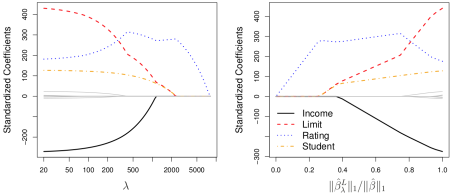
Ridge:

$$\sum_{j=1}^{p} \beta_j^2 \leq s$$

The Lasso:

$$\sum_{j=1}^{p} |\beta_j| \leq s$$





Linear transformation of predictors

Original Predictors:

$$X_1, \cdots, X_p$$

New Predictors:

$$Z_1, \cdots, X_M$$

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

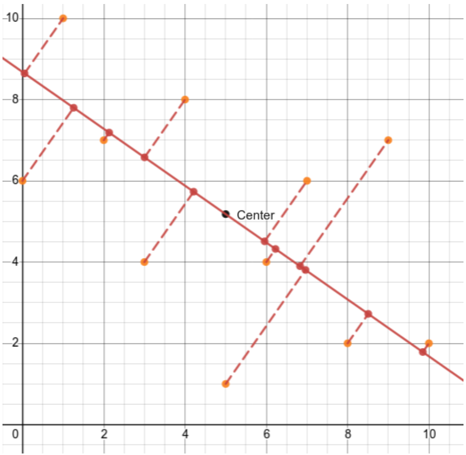
The goal:

- Find good φ's
- Fit regression model on Z_i's using least squares

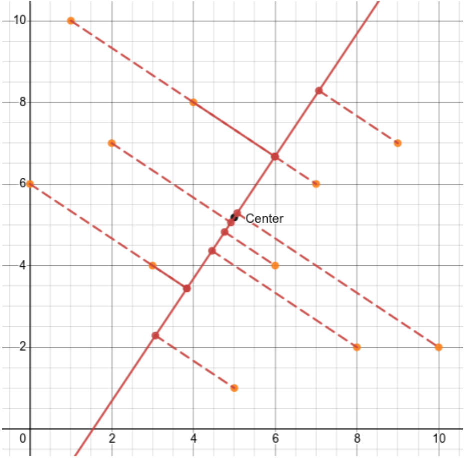
$$y_i = \theta_0 + \sum_{m=1}^{M} \theta_m z_{im} + \varepsilon_i$$

- Hope that lower dimensions means less overfitting
- Remember that interpretation not the same as shrinkage/subset selection of variables

PCA - First PC

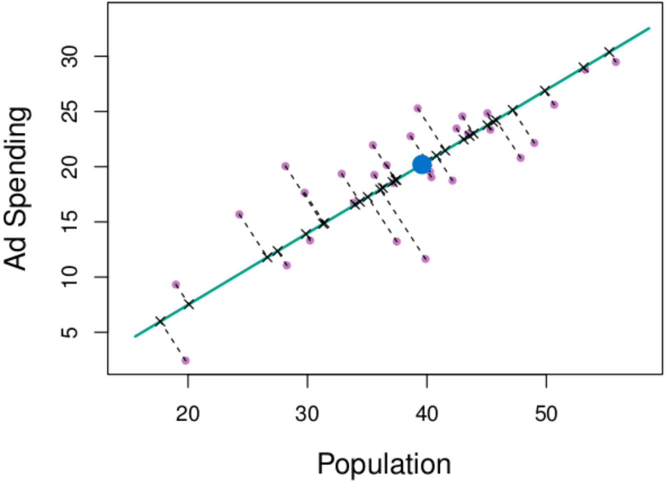


Projection has high variance



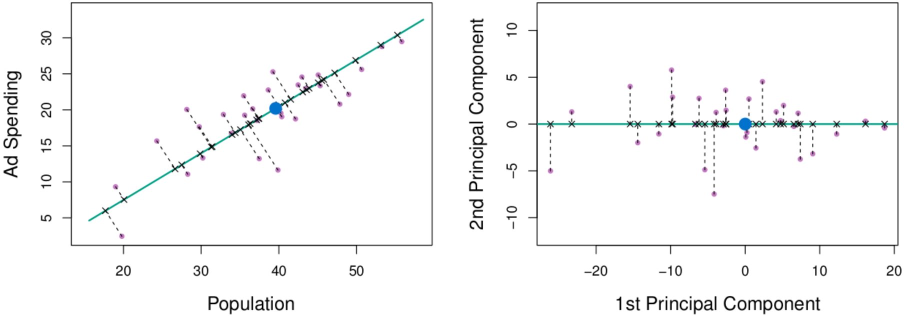
Projection has low(er) variance

Projection onto first PC



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

Drawing points in PC space



Principal Components Regression (PCR)

- Take new features Z_i
- Run regression
- Maybe do CV for a bunch of choices of M (where $M =$ number of Z_i features used) to pick a best M

Figuring out the original model from the PC model

- We have three input variables X_1, X_2, X_3
- We have two PCs,

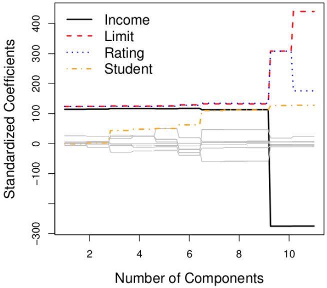
$$Z_1 = 0.266 \cdot X_1 - 0.077 \cdot X_2 + 0.961 \cdot X_3$$
$$Z_2 = 0.968 \cdot X_1 + 0.136 \cdot X_2 + -0.254 \cdot X_3$$

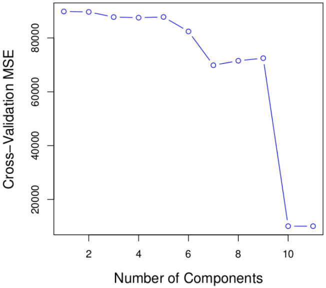
- Using linear regression, we learn the model

$$Y = -3 + 2Z_1 - 4Z_2.$$

- What are the coefficients for the model in terms of the X_i's?

Example on Credit dataset





Section 2

Partial Least Squares (PLS)

Supervised alternative

PCR: Non-supervised

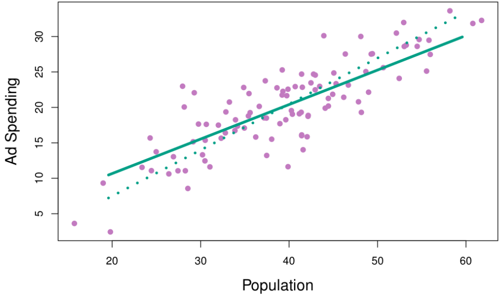
Partial Least Squares (PLS):

- Identify new features $Z_1, \cdots, Z_M$ linear combos of original where quality measure involves Y
- Fit linear model using least squares on these M features

First direction Z_1 for Partial Least Squares (PLS)

- Set φ_{j1} equal to the coefficient from simple linear regression of Y onto X_j
- The first direction is

$$Z_1 = \sum_{j=1}^{p} \varphi_{j1} X_j$$



Ex. Prediction of Y =Sales (not shown) on X_1 =Population and X_2 =Ad Spending

- Solid green: First PLS direction
- Dashed: First PC direction

Second (and more) PLS directions

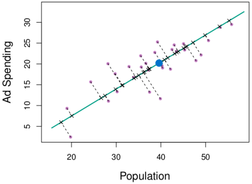
- Regress each variable on Z_1 and take residuals
- Compute Z_2 using *orthogonalized data* same as for Z_1

Code example on hitters data

PCA vs PLS

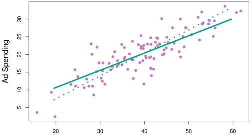
PCA

- 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



PLS

- Supervised dimensionality reduction
- Choose component Z_1 by using simple regression coefficients of each X_i onto Y
- Choose Z_2 and beyond by the same method after "getting rid" of info in the directions already explained
- Not a particular benefit, so usually default to PCA unless you have a good reason for this



Section 3

Issues in Higher Dimensions

High-Dimensional Data

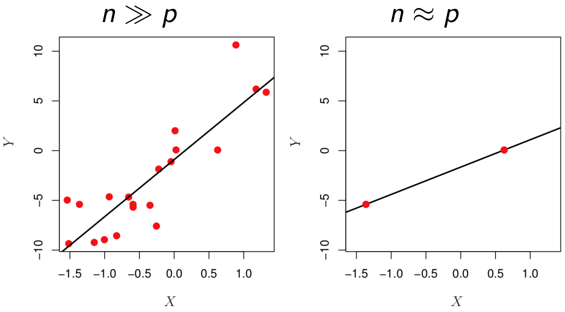
Low-Dimensions

$n \gg p$

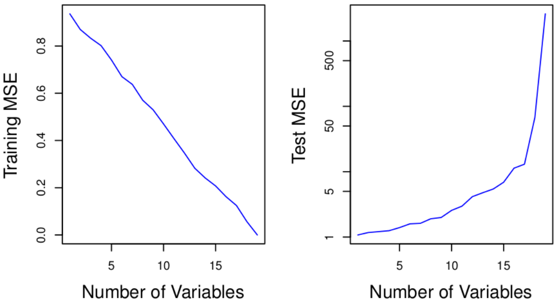
High-Dimensions

$n \ll p$

What goes wrong?



More issues with least squares on big p

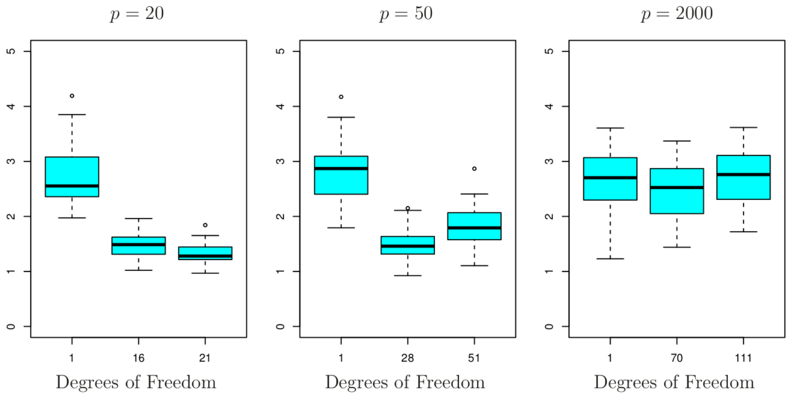


- $n = 20$
- Regression on $p = 1, \cdots, 20$
- Y completely unrelated to variables

The answer to dealing with big p

Be less flexible!

Example with Lasso



- $n = 100$
- Boxplots = Test MSE
- DF = # non-zero coeffs

Key points

Curse of dimensionality

Phenomena that arise when analyzing and organizing data in high-dimensional spaces that do not occur in low-dimensional settings.

Interpretation in high dimensions

Multi-collinearity: the concept that the variables in a regression might be correlated with each other

Reporting errors in high dimensions

Next time

10	M	Oct 3	Leave one out CV	5.1.1, 5.1.2	
11	W	Oct 5	k-fold CV	5.1.3	
12	F	Oct 7	More k-fold CV,	5.1.4-5	
13	M	Oct 10	k-fold CV for classification	5.1.5	HW #4 Due
14	W	Oct 12	Resampling methods: Bootstrap	5.2	
15	F	Oct 14	Subset selection	6.1	
16	M	Oct 17	Shrinkage: Ridge	6.2.1	HW #5 Due
17	W	Oct 19	Shrinkage: Lasso	6.2.2	
18	F	Oct 21	[No class, Dr Munch out of town]		
	M	Oct 24	No class - Fall break		
19	W	Oct 26	Dimension Reduction	6.3	
20	F	Oct 28	More dimension reduction; High dimensions	6.4	HW #6 Due
	M	Oct 31	Review		
	W	Nov 2	***Midterm #2***		