Ch 6.3, 6.4: PLS; Curse of Dimensionality

Lecture 21 - CMSE 381

Prof. Elizabeth Munch

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

Mon, Oct 30, 2023

Announcements

Last time:

- PCA/PCR

This lecture:

- Finish PCR
- 6.3: PLS
- 6.4: Issues with higher dimensions

Announcements:

- Homework #5 due Friday

Section 1

Last time

Shrinkage

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:

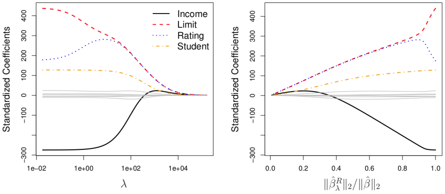
Least Squares:
No constraints

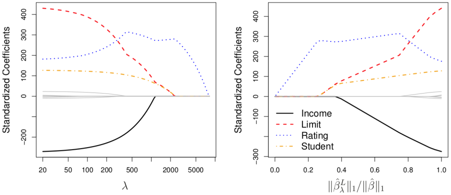
Ridge:

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

The Lasso:

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





Linear transformation of predictors

Original Predictors:

$$X_1, \cdots, X_p$$

New Predictors:

$$Z_1, \cdots, Z_M$$

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

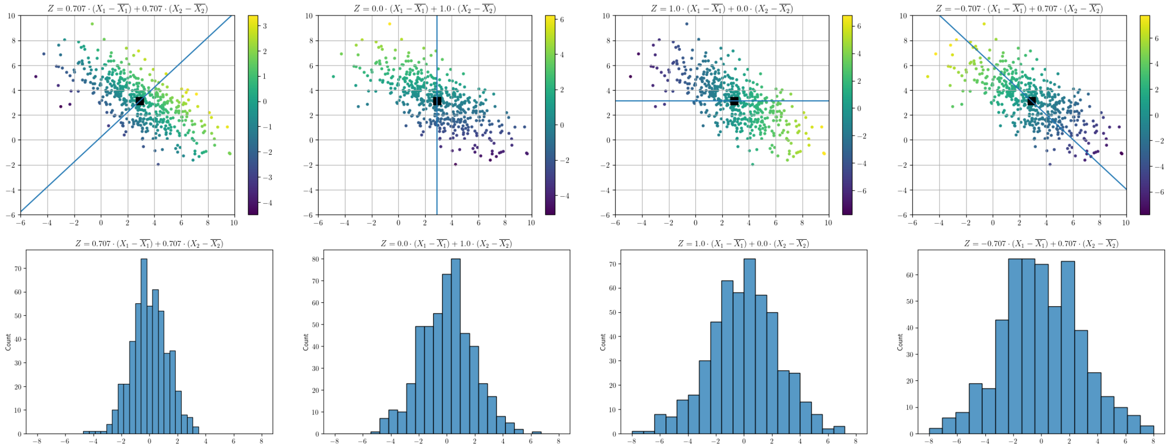
The goal:

- Find good φ's (PCA)
- Fit regression model on Z_i's using least squares (PLS)

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

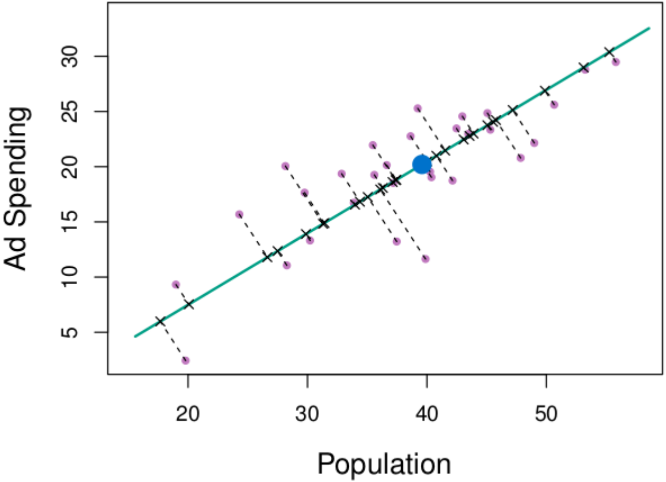
- Hope that lower dimensions means less overfitting

PCA - First PC



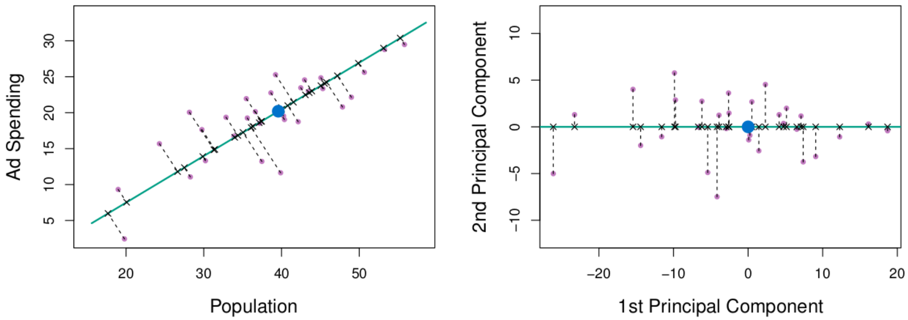
$Z = 0.707 \cdot (X_1 - \overline{X_1}) + 0.707 \cdot (X_2 - \overline{X_2})$
$Z = 0.0 \cdot (X_1 - \overline{X_1}) + 1.0 \cdot (X_2 - \overline{X_2})$
$Z = 1.0 \cdot (X_1 - \overline{X_1}) + 0.0 \cdot (X_2 - \overline{X_2})$
$Z = -0.707 \cdot (X_1 - \overline{X_1}) + 0.707 \cdot (X_2 - \overline{X_2})$
$Z = 0.707 \cdot (X_1 - \overline{X_1}) + 0.707 \cdot (X_2 - \overline{X_2})$
$Z = 0.0 \cdot (X_1 - \overline{X_1}) + 1.0 \cdot (X_2 - \overline{X_2})$
$Z = 1.0 \cdot (X_1 - \overline{X_1}) + 0.0 \cdot (X_2 - \overline{X_2})$
$Z = -0.707 \cdot (X_1 - \overline{X_1}) + 0.707 \cdot (X_2 - \overline{X_2})$
Count

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



Interpretation of PCR coefficients

Original Predictors:

$$X_1, \cdots, X_p$$

New Predictors:

$$Z_1, \cdots, Z_M$$

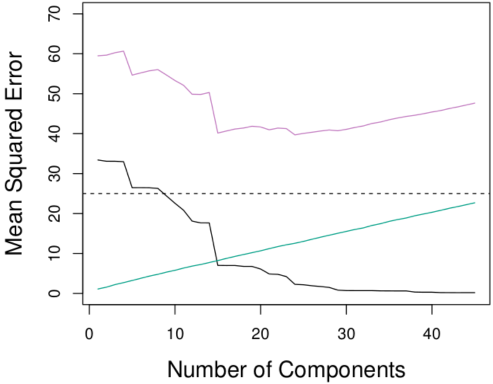
$$Z_m = \sum_{j=1}^{p} \varphi_{jm} X_j$$

Learned model:

$$y = \theta_0 + \theta_1 Z_1 + \cdots + \theta_M Z_M$$

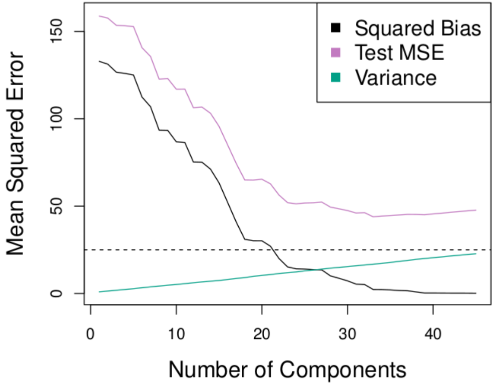
Bias variance trade-off

Example with simulated data: $n = 50$ observations of $p = 45$ predictors
Y is a function of **all predictors**



Bias variance trade-off

Example with simulated data: $n = 50$ observations of $p = 45$ predictors
Y is a function of **2 predictors**



Properties of PCR

Section 2

Partial Least Squares (PLS)

Supervised alternative

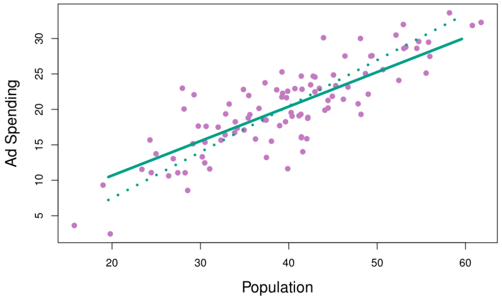
PCR: Non-supervised

Partial Least Squares (PLS):

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

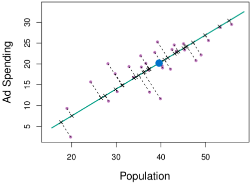
Second (and more) PLS directions

- 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

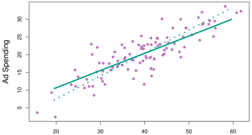
PCR

- Unsupervised dimensionality reduction + linear regression
- 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

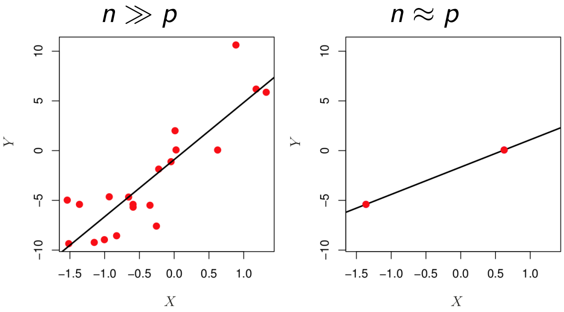
Low-Dimensions

$n \gg p$

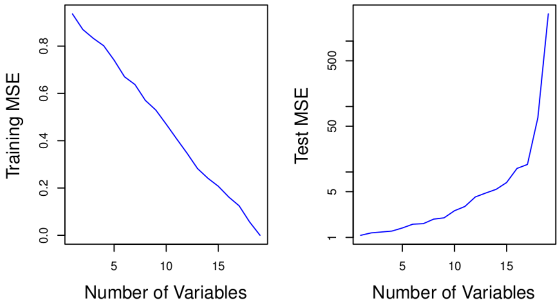
High-Dimensions

$n \ll p$

What goes wrong?



More issues with least squares on big p

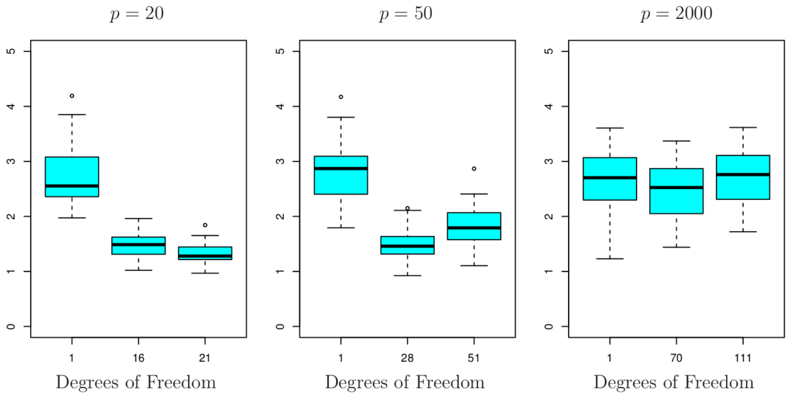


- $n = 20$
- Regression on $p = 1, \cdots, 20$
- Y completely unrelated to variables

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

Lec #	Date			Reading	Homeworks
20	Fri	Oct 27	Dimension Reduction	6.3	
21	Mon	Oct 30	More dimension reduction; High dimensions	6.4	
22	Wed	Nov 1	Polynomial & Step Functions	7.1,7.2	
23	Fri	Nov 3	Step Functions	7.2	HW #6 Due
24	Mon	Nov 6	Basis functions, Regression Splines	7.3,7.4	
25	Wed	Nov 8	Decision Trees	8.1	
26	Fri	Nov 10	Random Forests	8.2.1, 8.2.2	
27	Mon	Nov 13	Maximal Margin Classifier	9.1	
28	Wed	Nov 15	SVC	9.2	
29	Fri	Nov 17	SVM	9.3, 9.4	
30	Mon	Nov 20	Single layer NN	10.1	
31	Wed	Nov 22	Overflow/project day?		
	Fri	Nov 24	No class - Thanksgiving		
	Mon	Nov 27	**Review**		
	Wed	Nov 29	**Midterm #3**		
32	Fri	Dec 1	Multi Layer NN	10.2	
33	Mon	Dec 4	CNN	10.3	
34	Wed	Dec 6	Unsupervised Learning & Clustering	12.1, 12.4	
35	Fri	Dec 8	Overflow/Project day?		Project due