

Per Bruun Brockhoff

DTU Compute, Statistics and Data Analysis  
 Building 324, Room 220  
 Danish Technical University  
 2800 Lyngby – Danmark  
 e-mail: perbb@dtu.dk

Overview

- 1 Multivariate Calibration
  - Optimal predictor
- 2 Example data
- 3 Ridge regression
- 4 Lasso regression
- 5 Ridge regression: Solution
- 6 The  $L_2$ -methods
- 7 All methods
- 8 Little Ridge example
- 9 Ridge/Lasso - How to do it! (same as for PCR/PLS)

Multivariate calibration

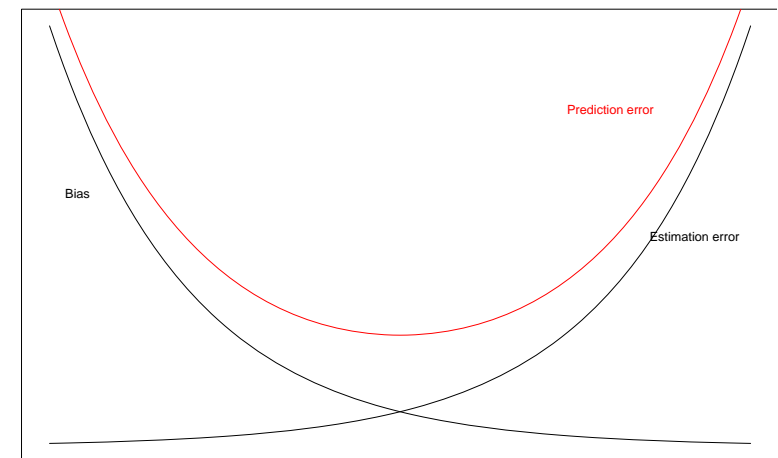
- Classical (chemical) calibration: Model  $x$  as function of  $y$
- Predict  $y$  fro  $x$  by "going backwards".
- We are doing Inverse calibration:  $y = f(x)$
- What is a good model: In theory:  
 $f(x) = E(Y|X = x)$ .
- If  $Y = E(Y|X = x) + \varepsilon$  and  $\hat{f}(x)$  is some model estimator, expected prediction error:

$$E_{Y|X=x}((y - \hat{f}(x))^2) =$$

$$E((y - E(Y|X = x))^2) + E((f(x) - \hat{f}(x))^2) + (f(x) - E(Y|X = x))^2$$

$$= \sigma^2 + \text{Var}(\hat{f}(x)) + \text{Bias}^2$$

Find optimal predictor by CV



## Basic model: MLR

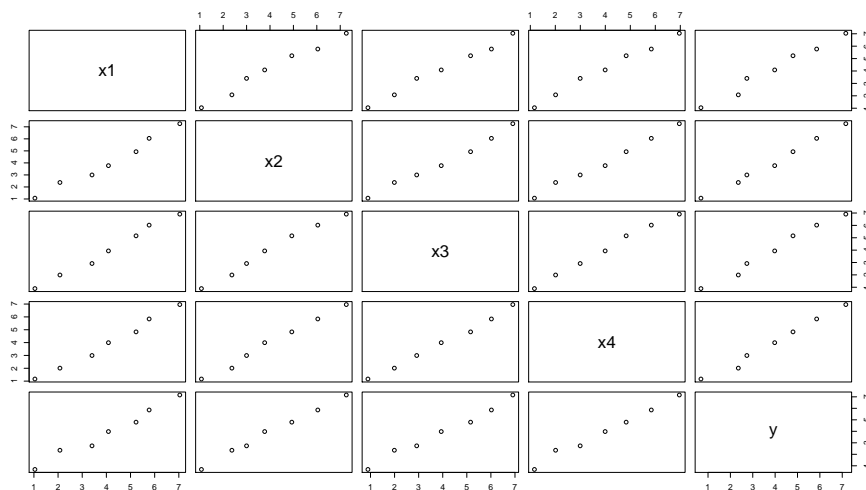
$$E(y_i|x_{i1}, x_{i2}, \dots, x_{ip}) = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip}$$

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$$

- Estimation:  $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
- Problem: The parameters are badly estimated!
- HIGH Uncertainty!
- HIGH value of one parameter can go together with a parameter of opposite sign - NOT changing the model

## Example data

## Example Data



## Basic model: MLR

- Uncertainty:

$$\text{Var}(\hat{\mathbf{b}}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

$$= \hat{\sigma}^2 \sum_{j=1}^p (\mathbf{v}_j\mathbf{v}_j')/d_j^2$$

$\mathbf{v}_j$  is the  $j$ th principal loading vector  
 $d_j^2$  is the  $j$ th principal eigenvalue of  $\mathbf{X}'\mathbf{X}$   
 $(\text{Var}(\mathbf{X}\mathbf{v}_j) = d_j^2/n)$

- SMALL principal directions will induce HIGH uncertainty!

## Example data

## MLR results:

	Estimate	Std. Error	t	P-val
Intercept	-0.1668	0.4309	-0.387	0.736
x1	-0.8141	1.6888	-0.482	0.677
x2	-0.1027	0.8635	-0.119	0.916
x3	2.0695	1.4133	1.464	0.281
x4	-0.1354	0.6183	-0.219	0.847

$b_3$  becomes too large AND the others too small.

## Ridge regression

- Penalizes (numerically) large regression coefficients
- $\hat{\mathbf{b}}^{Ridge}$  minimizes (for a given  $s$ )

$$\sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_{ij} \right)^2 \quad \text{subject to} \quad \sum_{j=1}^p b_j^2 \leq s$$

- $\hat{\mathbf{b}}^{Ridge}$  minimizes (for a given  $\lambda$ )

$$\sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p b_j^2$$

DTU Compute  
Department of Applied Mathematics and Computer Science

## Lasso regression

- Penalizes (numerically) large regression coefficients
- $\hat{\mathbf{b}}^{Lasso}$  minimizes (for a given  $s$ )

$$\sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_{ij} \right)^2 \quad \text{subject to} \quad \sum_{j=1}^p |b_j| \leq s$$

- $\hat{\mathbf{b}}^{Lasso}$  minimizes (for a given  $\lambda$ )

$$\sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |b_j|$$

DTU Compute  
Department of Applied Mathematics and Computer Science

## Ridge regression: Solution

- Add  $\lambda$  to the diagonal of  $\mathbf{X}'\mathbf{X}$ :

$$\mathbf{b}^{Ridge} = (\mathbf{X}'\mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}'\mathbf{y}$$

- RR-predictor:  $\hat{\mathbf{y}}^{Ridge} = \mathbf{X}\mathbf{b}^{Ridge}$
- Operational: Try different  $\lambda$ s - choose the optimal by Cross Validation.

DTU Compute  
Department of Applied Mathematics and Computer Science

## MLR(LS), PCR and RR

$$\hat{\mathbf{y}}^{LS} = \mathbf{X}\mathbf{b}^{LS} = \sum_{j=1}^p \mathbf{u}_j \mathbf{u}_j' \mathbf{y}$$

$$\hat{\mathbf{y}}^{Ridge} = \mathbf{X}\mathbf{b}^{Ridge} = \sum_{j=1}^p \mathbf{u}_j \frac{d_j^2}{d_j^2 + \lambda} \mathbf{u}_j' \mathbf{y}$$

$$\hat{\mathbf{y}}^{PCR} = \mathbf{U}\boldsymbol{\theta}^{PCR} = \mathbf{X}\mathbf{b}^{PCR} = \sum_{j=1}^A \mathbf{u}_j \mathbf{u}_j' \mathbf{y}$$

$\mathbf{u}_j$  is the  $j$ th principal score values

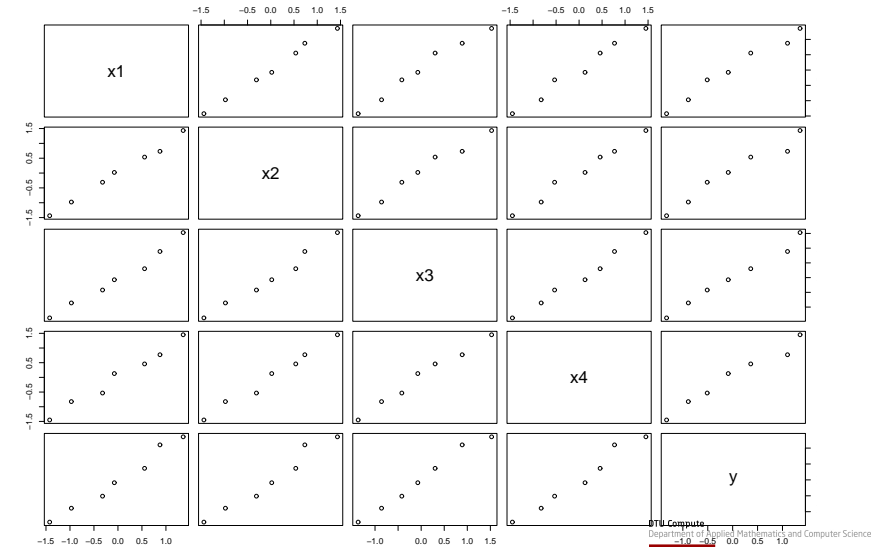
DTU Compute  
Department of Applied Mathematics and Computer Science

## MLR(LS), PCR and RR/Lasso

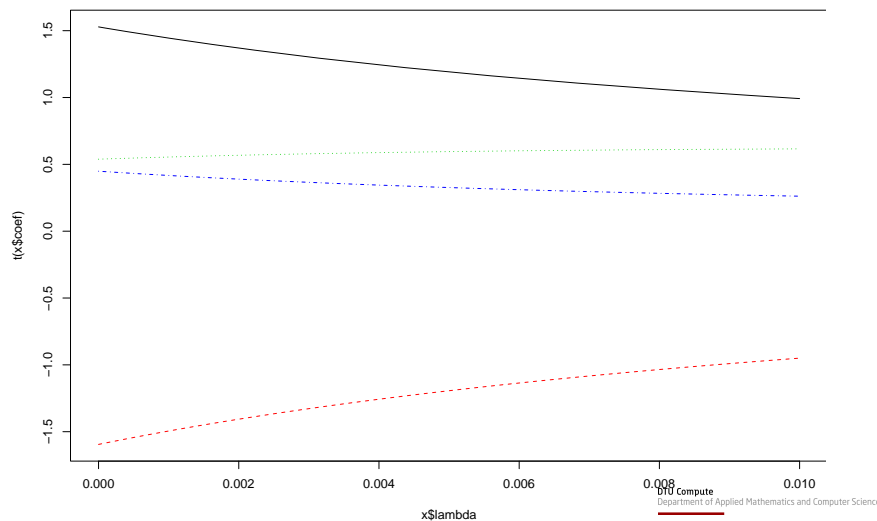
- The LS-solution is a vector in space.
- Ridge regression *shrinks* the LS-solution
- Highest shrinkage in low-variance directions
- PCR also shrinks: No shrinkage in main directions, 100% shrinkage in low variance directions.
- (PLS is ALSO a shrinkage method - although "expansion" is also possible)
- Lasso also shrinks - some coefficients down to EXACTLY zero
- Lasso is hence also an automated variable selection method
- Lasso is an  $L_1$  method - VERY different from:
- The others are  $L_2$  methods

DTU Compute  
Department of Applied Mathematics and Computer Science

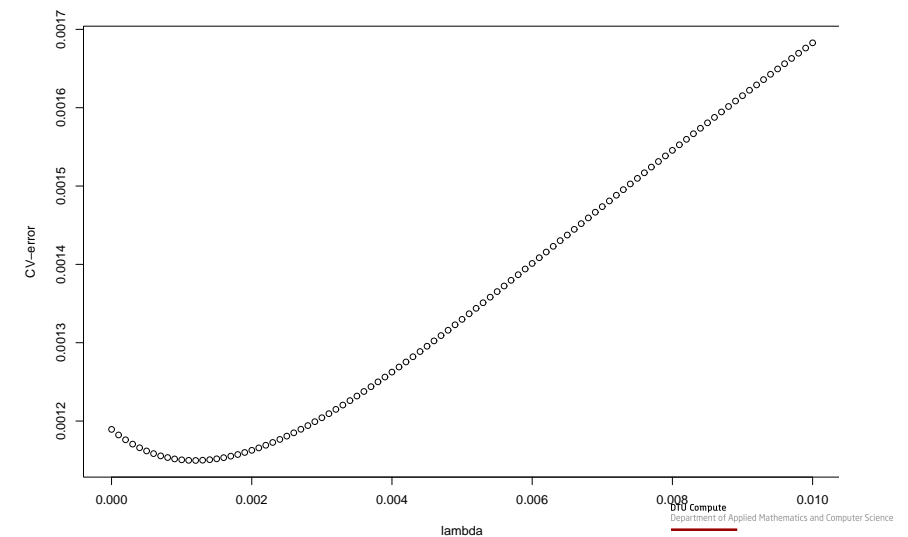
## Example Data

DTU Compute  
Department of Applied Mathematics and Computer Science

## Regression coefficients

DTU Compute  
Department of Applied Mathematics and Computer Science

## Cross Validation

DTU Compute  
Department of Applied Mathematics and Computer Science

## MLR and RR results:

	MLR Estimate	RR estimate ( $\lambda = 0.0012$ )
x1	1.651	1.543
x2	-1.723	-1.595
x3	0.5819	0.6023
x4	0.4848	0.4440

## How to do it?(same as for PCR/PLS)

- 1 Explore data
- 2 Do modelling (choose number of components, consider variable selection)
- 3 Validate (residuals, outliers, influence etc) (not so straightforward)
- 4 Iterate e.g. on 2. and 3.
- 5 Interpret, conclude, report.
- 6 If relevant: predict future values.