Penalised Regression

NGSchool2022

Nuno Sepúlveda, 16.09.2022
nuno.Sepulveda@mini.pw.edu.pl
http://www.immune-stats.net

Objectives

Basics on penalised regression

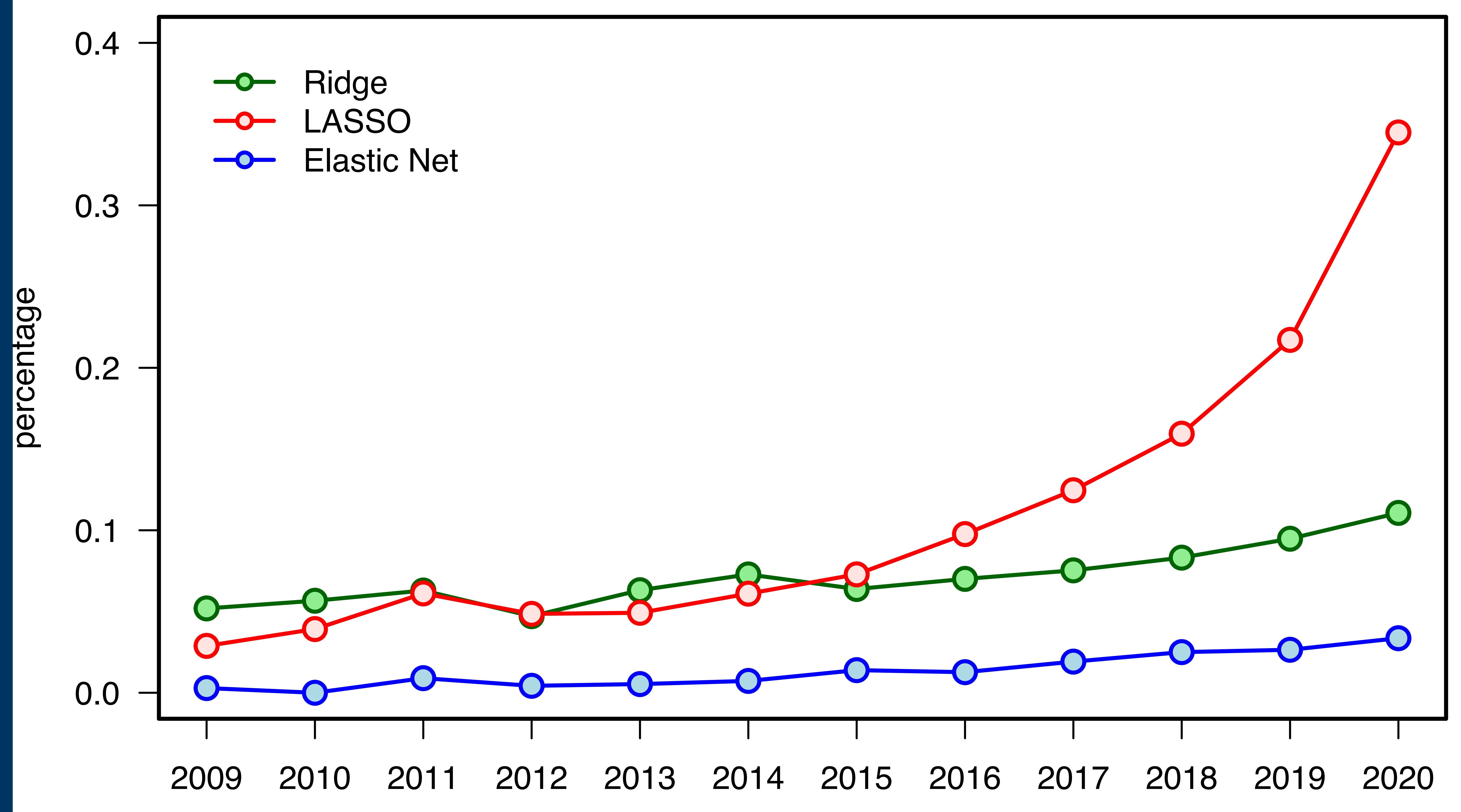
Ridge Regression

LASSO Regression

Elastic Net Regression

Use R to conduct data analysis

Why learning linear regression?



Estimation

Model selection


Accuracy
Bias

Penalised regression

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_i \right)^2 \right\} .$$

subject to a constraint

$$pen \leq \lambda \ .$$

Ridge Regression

$$\hat{\boldsymbol{b}} = \underset{\boldsymbol{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_i \right)^2 \right\} ,$$

subject to $\displaystyle\sum_{j=1}^{p} b_j^2 \leq \lambda_2$ $\qquad$ $\displaystyle\lambda_2 \in \left[0, \sum_{j=1}^{p} (\hat{b}_j^*)^2 \right]$

Hoerl AE, Kennard RW (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12(1): 55-67.

Geometrical interpretation (2D)

$$\sum_{j=1}^{2} b_j^2 \leq \lambda_2$$

$$b_1 = r\cos\theta$$

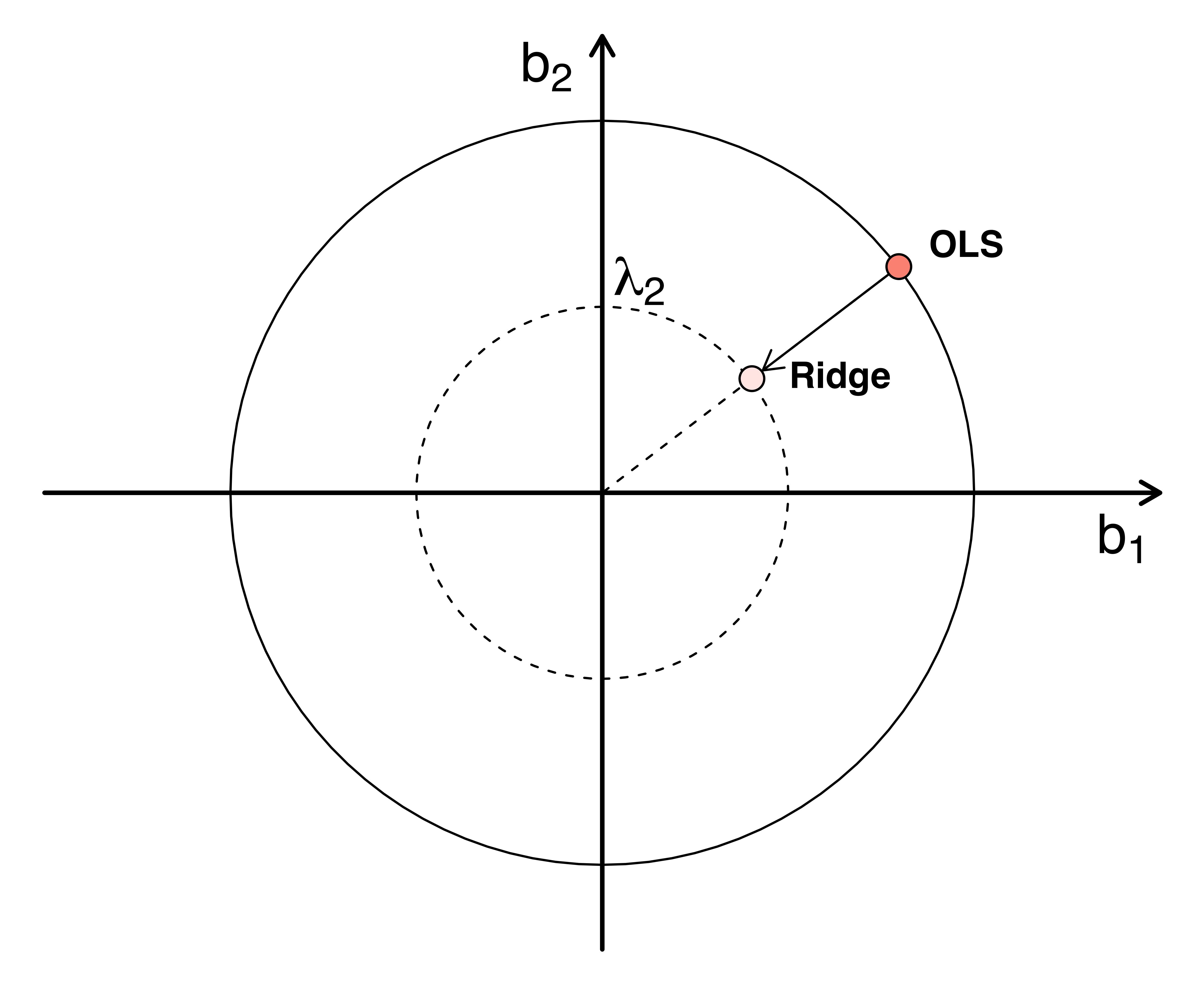
$$b_2 = r\sin\theta$$

$$r^2(\cos^2\theta + \sin^2\theta) \leq \lambda_2$$

$$r^2 \leq \lambda_2$$

Ridge estimator is only dependent on the radius and not on the angle

Geometrical interpretation (2D)



Ordinary least squares estimator

$$\hat{\mathbf{b}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$

Ridge estimator

$$\hat{\mathbf{b}} = \left(\mathbf{X}^T\mathbf{X} + \lambda_2\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$

Ridge Regression

$$\hat{\boldsymbol{b}} = \underset{\boldsymbol{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_i \right)^2 \right\} ,$$

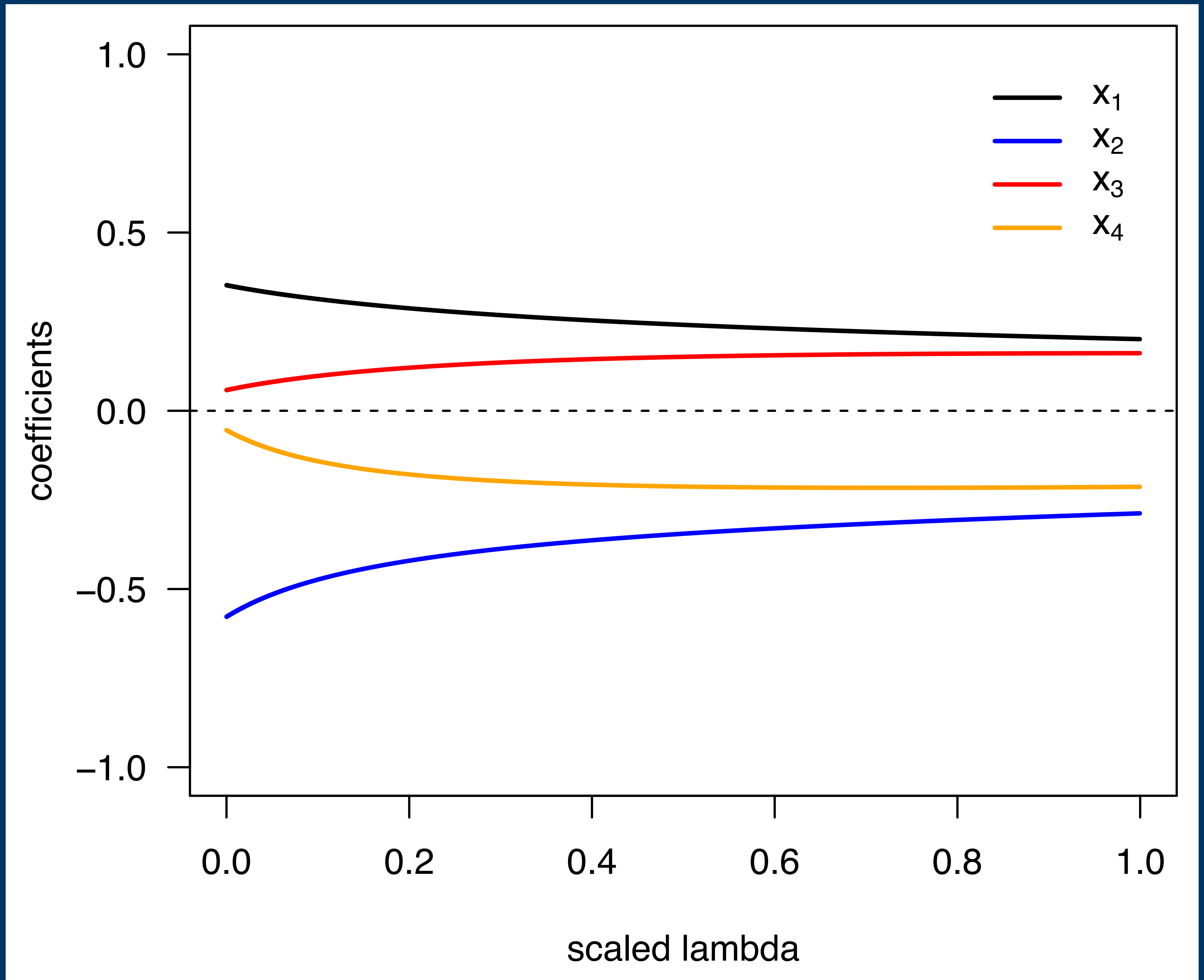
subject to $$\frac{\sum_{j=1}^{p} b_j^2}{\sum_{j=1}^{p} (\hat{b}_j^*)^2} \leq 1 - \lambda^*$$

0% shrinkage

$$\lambda^* \in [0,1]$$

“100%” shrinkage

Example



1.0
0.5
0.0
−0.5
−1.0
coefficients
0.0
0.2
0.4
0.6
0.8
1.0
scaled lambda
x_1
x_2
x_3
x_4

Ridge regression

Advantages	Disadvantages
Remove multicollinearity	Biased estimators
Estimator with a closed form	No shrinkage to zero
Shrinkage	(No model selection)

LASSO Regression

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_i \right)^2 \right\} ,$$

subject to a constraint

$$\sum_{j=1}^{p} |\, b_j \,| \leq \lambda_1 \qquad \lambda_1 \in \left[0, \sum_{j=1}^{p} |\, \hat{b}_j^* \,| \right]$$

Geometrical interpretation (2D)

$$\sum_{j=1}^{2} |b_j| \leq \lambda_1$$

$$r(\cos\theta + \sin\theta) \leq \lambda_2$$

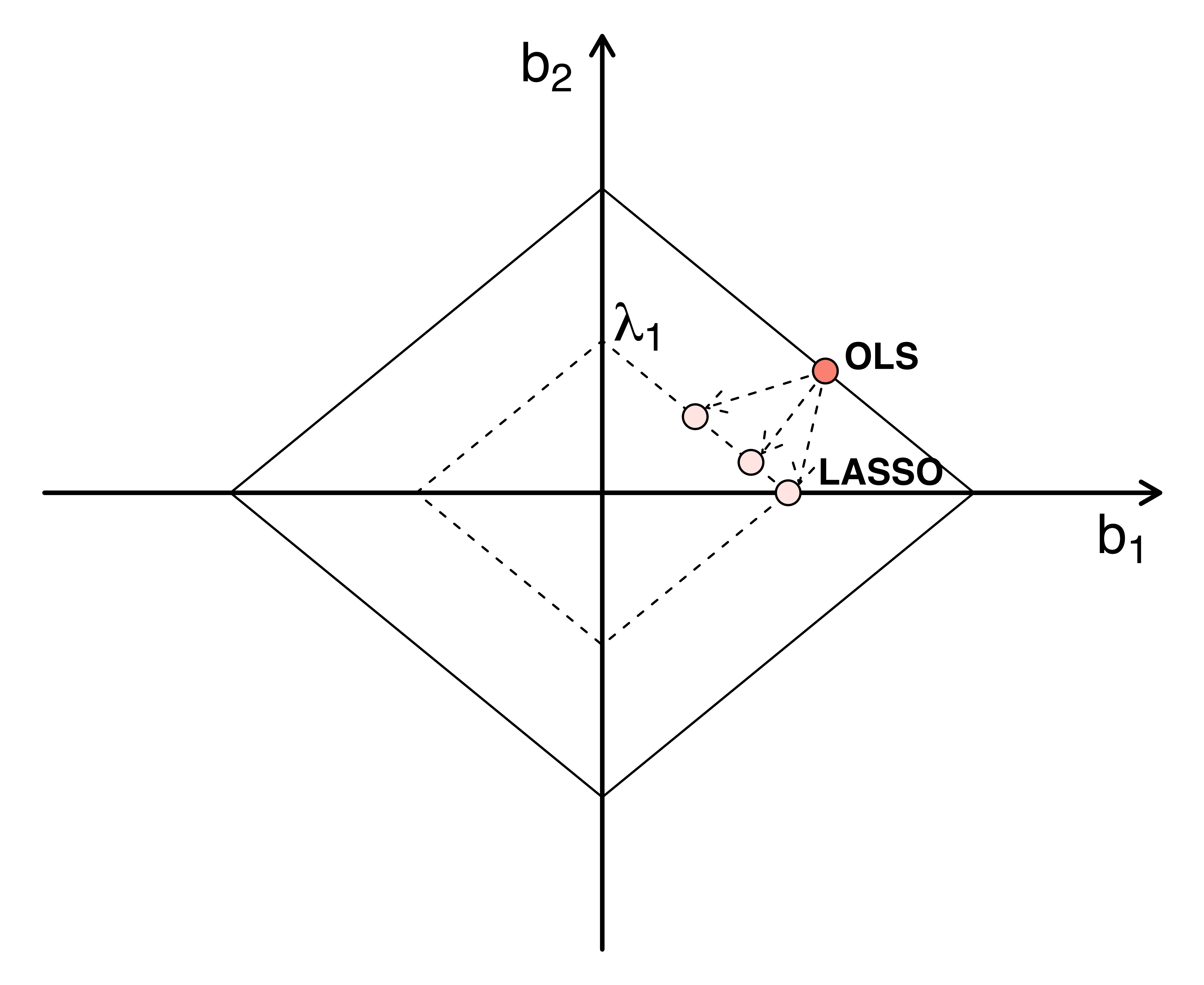
$$r^2 \leq \lambda_2$$

$$b_1 = r\cos\theta$$

$$b_2 = r\sin\theta$$

LASSO estimator is dependent on both radius and angle

Geometrical interpretation (2D)



LASSO Regression

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_i \right)^2 \right\} ,$$

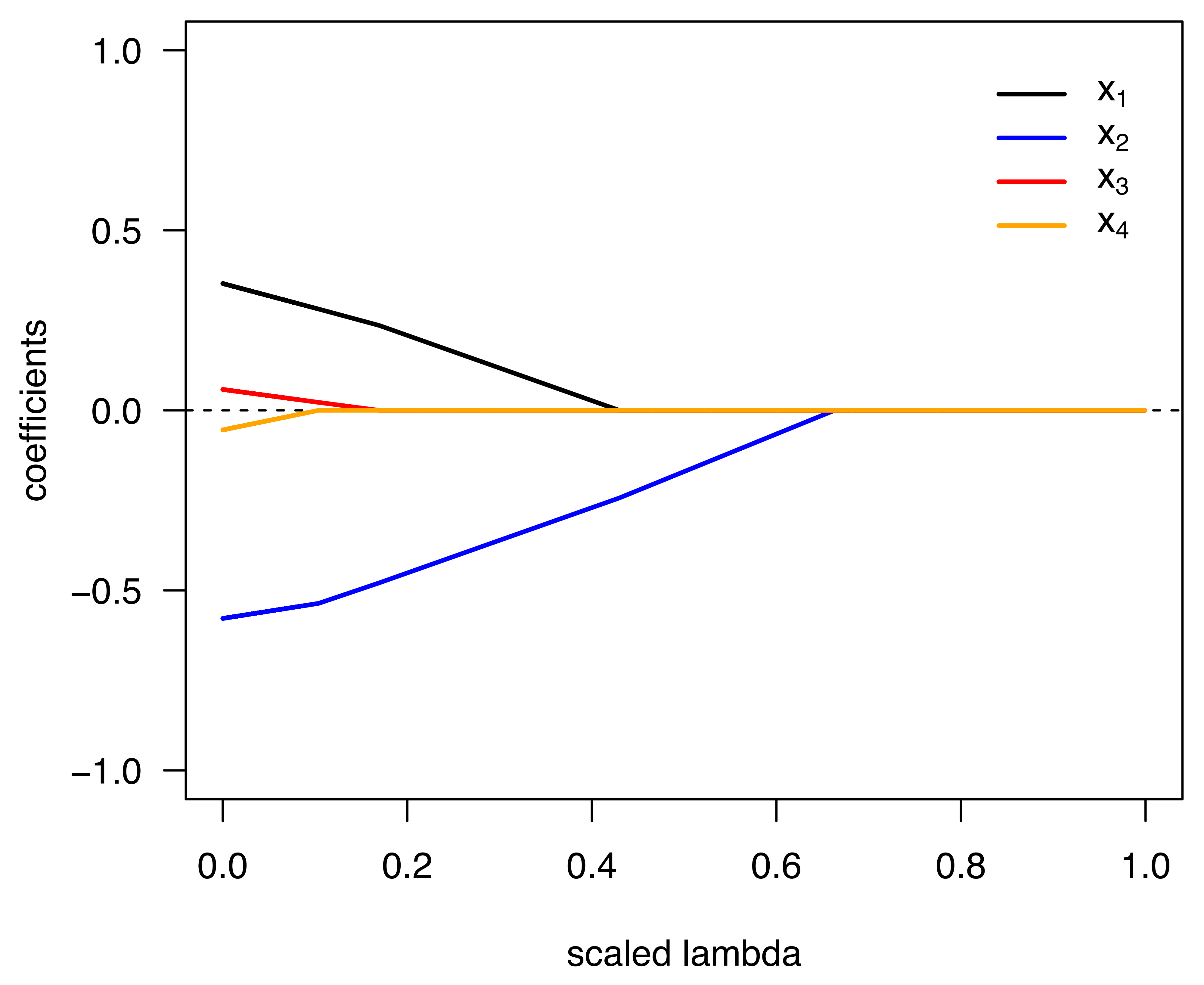
0% shrinkage

subject to $$\frac{\sum_{j=1}^{p} |b_j|}{\sum_{j=1}^{p} |b_j^*|} \le 1 - \lambda^*$$

$$\lambda^* \in [0,1]$$

100% shrinkage

Example



1.0
0.5
0.0
−0.5
−1.0
coefficients
0.0
0.2
0.4
0.6
0.8
1.0
scaled lambda
x_1
x_2
x_3
x_4

LASSO regression

Advantages	Disadvantages
Remove multicollinearity	Random choice of highly correlated covariates
Shrinkage to zero	No closed-form expression
(Model selection)	Problems with the standard errors

Elastic Net Regression

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_i \right)^2 \right\} ,$$

subject to $\alpha \,|\, \mathbf{b} \,|_1 + (1-\alpha) \,|\, \mathbf{b} \,|^2 \leq \lambda$ for some $\lambda, \alpha \in [0,1]$.

$\alpha = 0 \Rightarrow$ Ridge Regression

$\alpha = 1 \Rightarrow$ LASSO Regression

Estimation of the tuning parameter(s)

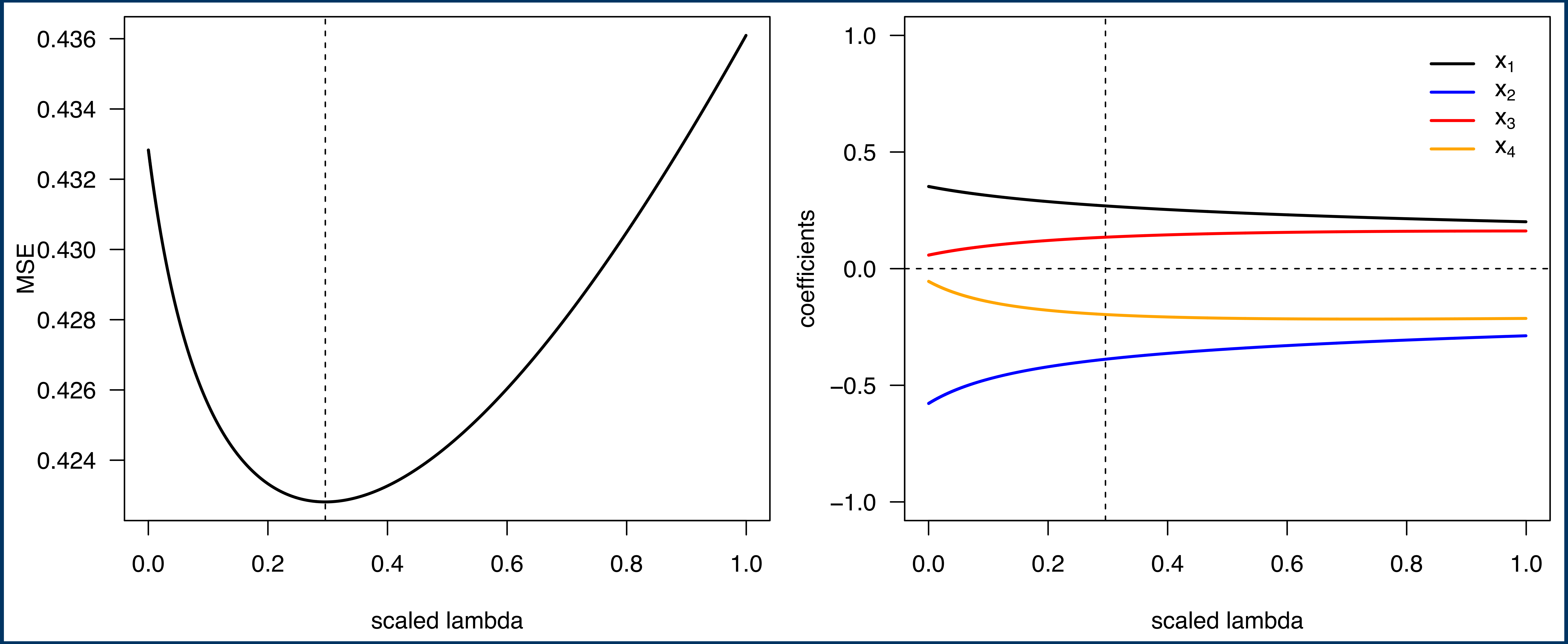
Evaluate a grid of possible values

Highest accuracy

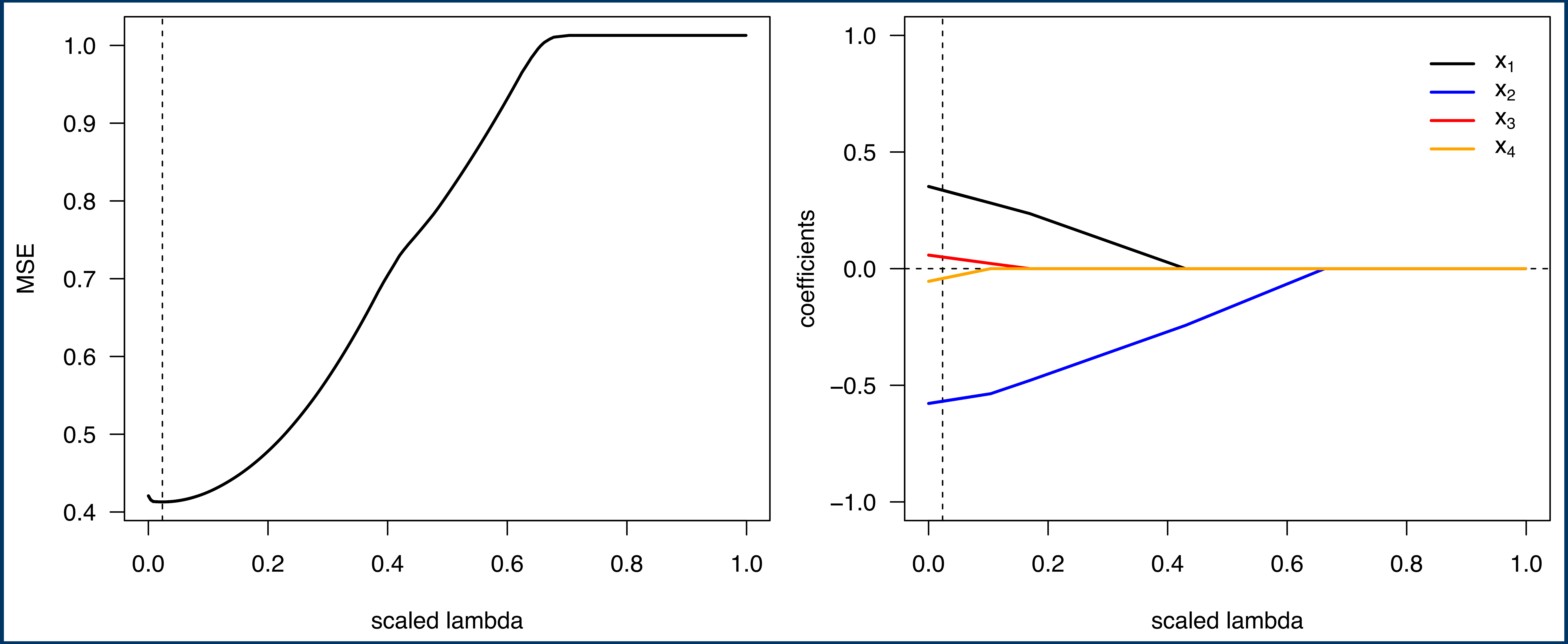
Cross-validation

Lowest mean squared error

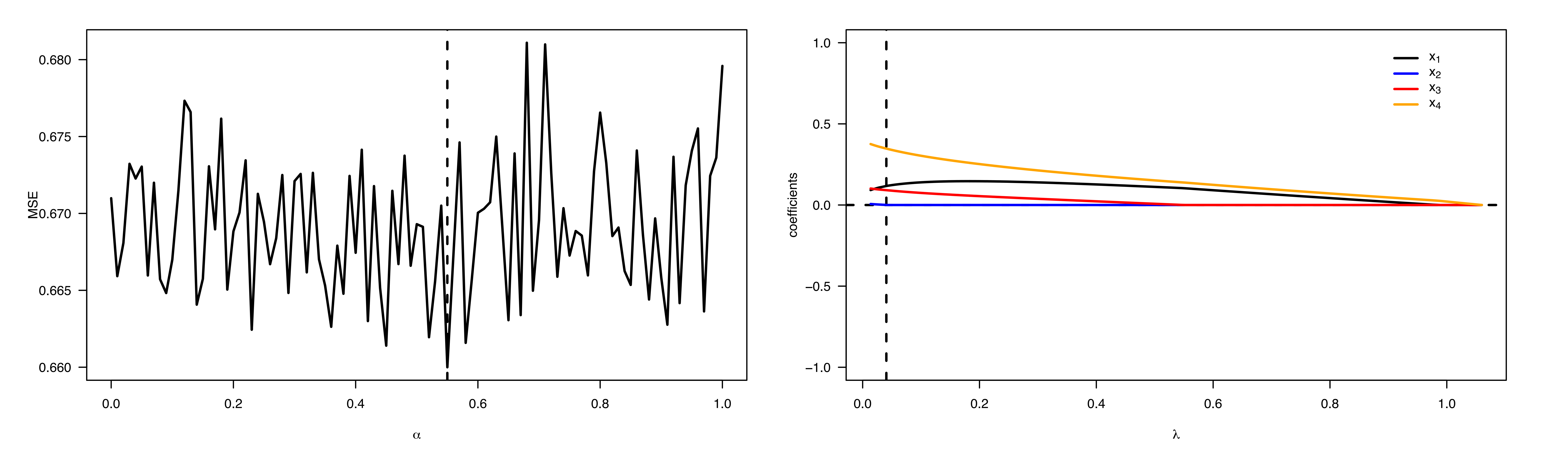
Example: Ridge Regression



Example: LASSO Regression



Example: Elastic Net Regression



Other penalised regression methods

- Adaptive LASSO
- Smoothly clipped Absolute Deviations (SCAD)
- Minimax Concave Penalty (MCP)
- Bayesian LASSO

Acknowledgements

Immune-Stats Group
André Fonseca

MiNI@WUT
Przemyslaw Biecek

NGSchool 2022
Katarzyna Kedzierska
Urszula Smyczynska
Gladys Wojciechowska

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

ref. UIDB/00006/2020

NAWA
POLISH NATIONAL AGENCY FOR ACADEMIC EXCHANGE

ref. PPN/ULM/2020/1/00069/U/00001