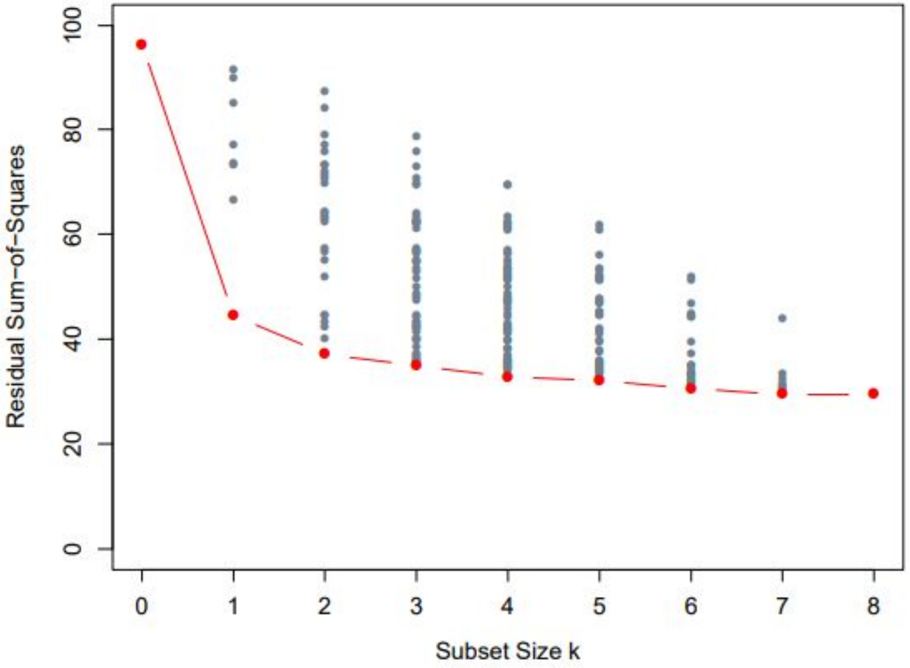
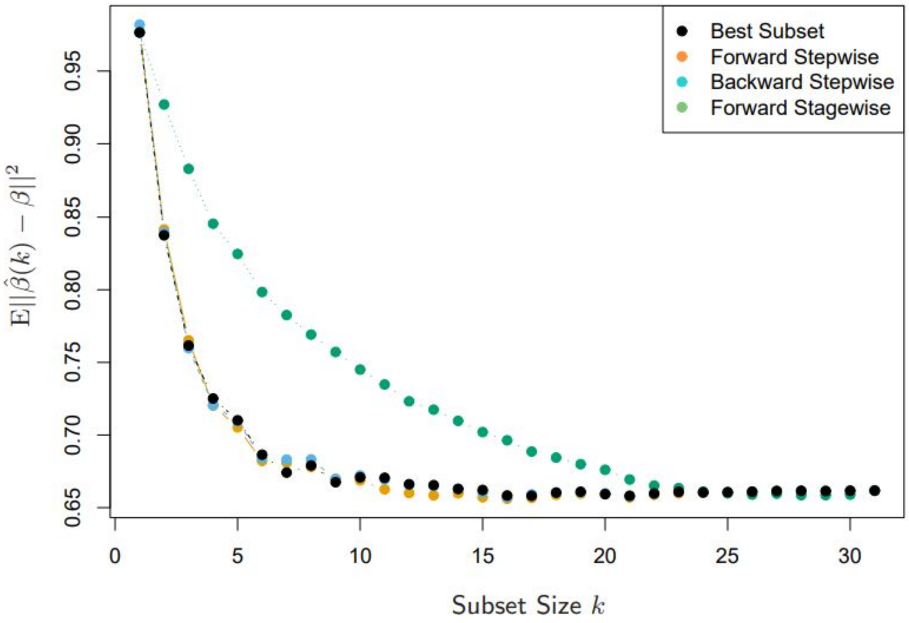
Wybór podzbioru

Best-Subset Selection



$$k \in \{0, 1, 2, \ldots, p\}$$

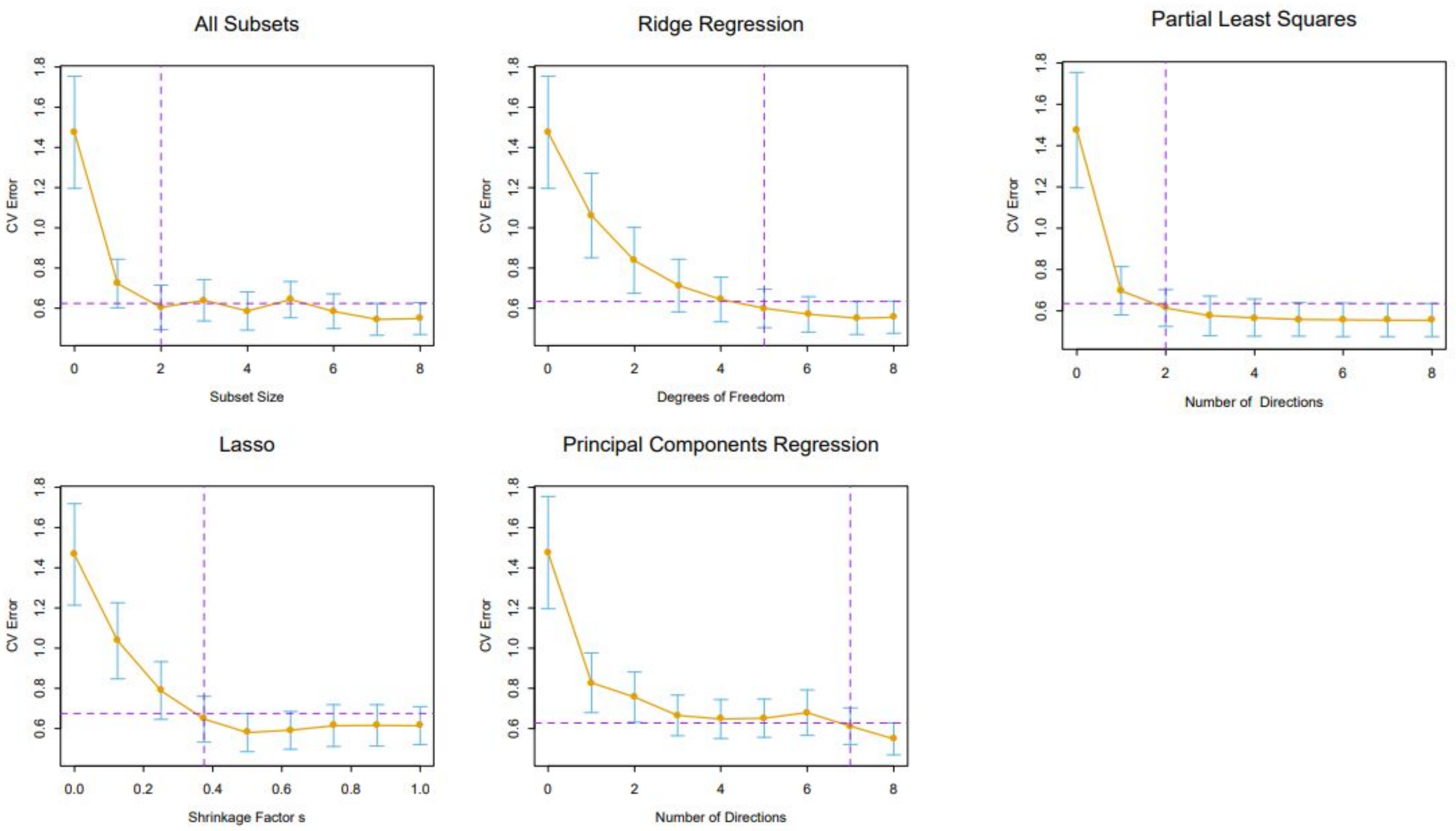
Forward- and Backward-Stepwise Selection



$$N = 300$$

$$p = 31$$

Shrinkage Methods



All Subsets
Subset Size
Ridge Regression
Degrees of Freedom
Partial Least Squares
Number of Directions
Lasso
Shrinkage Factor s
Principal Components Regression
Number of Directions
CV Error

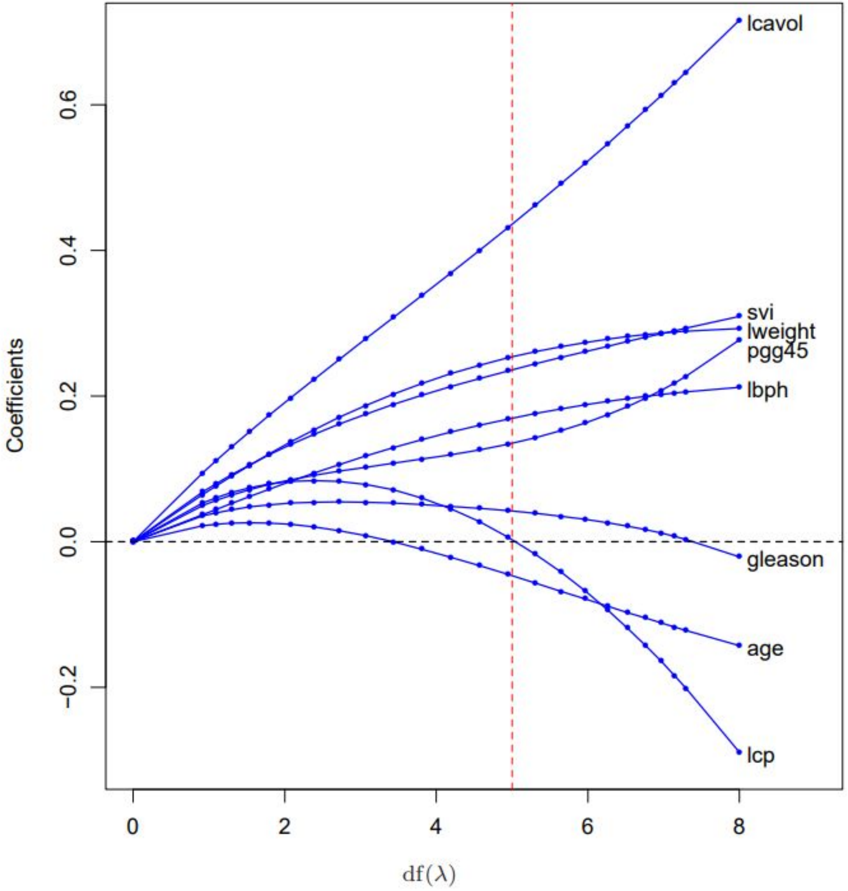
Ridge Regression

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\text{argmin}} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}$$

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2$$ przy założeniu $$\sum_{j=1}^{p} \beta_j^2 \le t$$

estymujemy β_0 przez $\bar{y} = \frac{1}{N} \sum_1^N y_i$

$$\text{RSS}(\lambda) = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda \beta^T \beta \qquad \hat{\beta}^{\text{ridge}} = (\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$



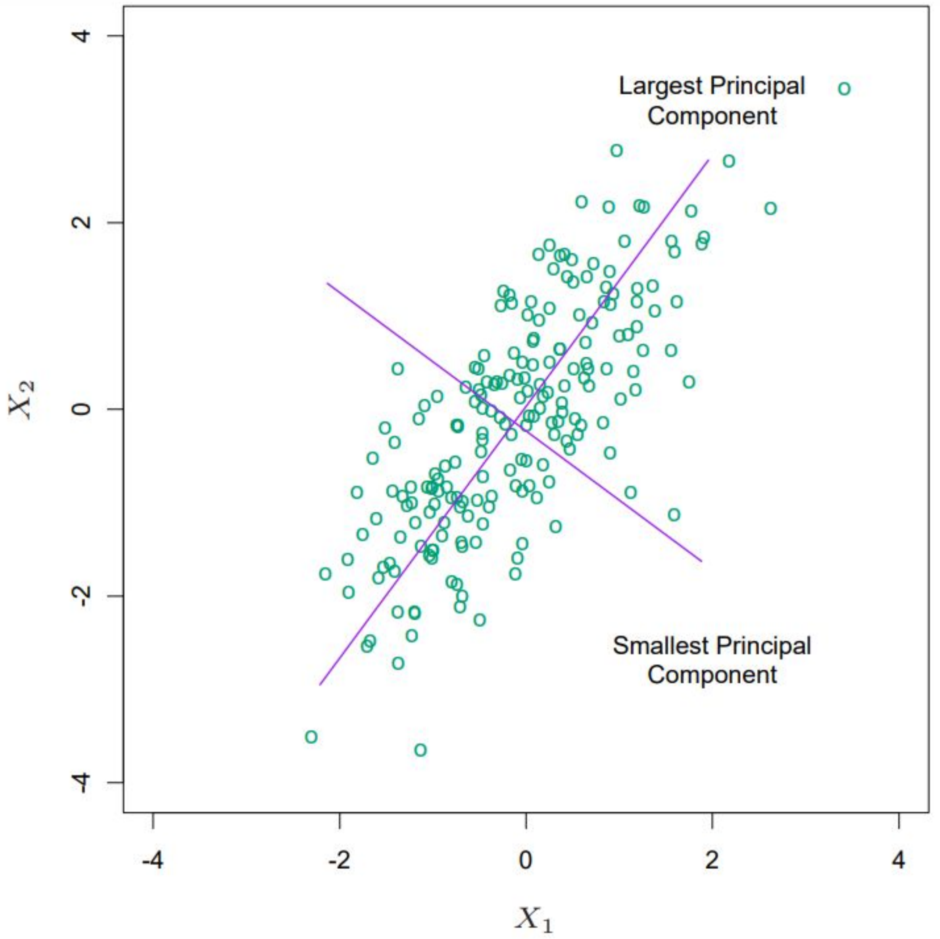
$$\hat{\beta}^{\text{ridge}} = \hat{\beta}/(1+\lambda)$$

SVD macierzy X o wymiarach Nxp

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

$$\mathbf{X}\hat{\beta}^{\text{ls}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{U}\mathbf{U}^T\mathbf{y}$$

$$\begin{aligned}
\mathbf{X}\hat{\beta}^{\text{ridge}} &= \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \\
&= \mathbf{U}\,\mathbf{D}(\mathbf{D}^2 + \lambda\mathbf{I})^{-1}\mathbf{D}\,\mathbf{U}^T\mathbf{y} \\
&= \sum_{j=1}^{p} \mathbf{u}_j \frac{d_j^2}{d_j^2 + \lambda} \mathbf{u}_j^T\mathbf{y},
\end{aligned}$$



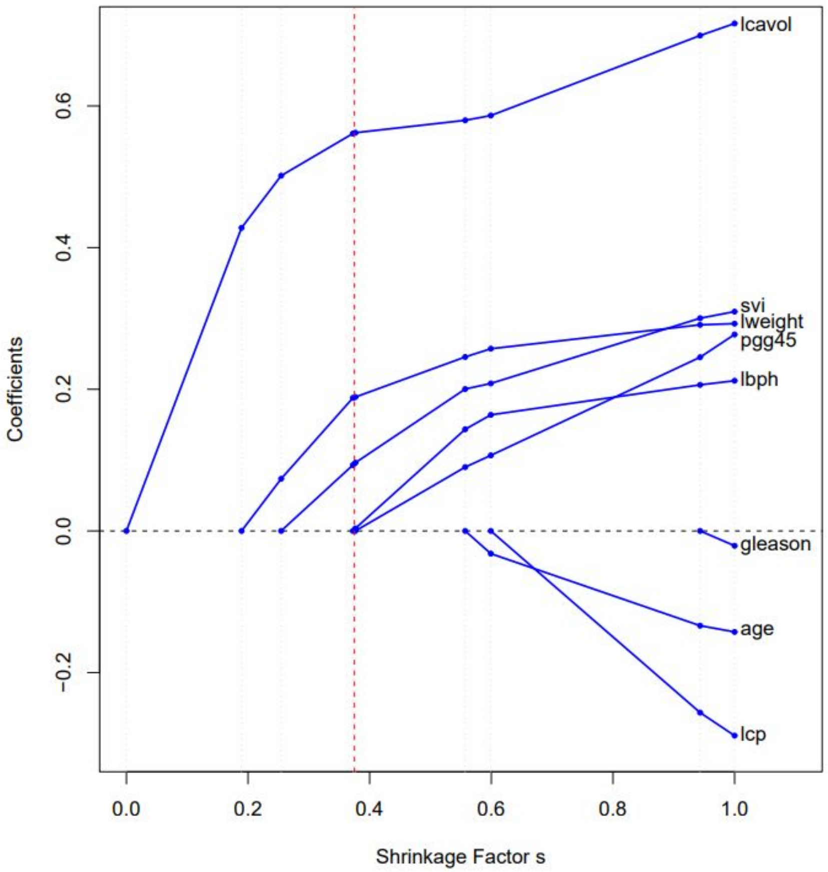
Largest Principal Component
Smallest Principal Component
4
2
0
-2
-4
X_2
-4
-2
0
2
4
X_1

Lasso

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2$$

przy założeniu $\sum_{j=1}^{p} |\beta_j| \leq t$

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$



$$s = t / \sum_1^p |\hat{\beta}_j|$$

Porównanie: wybór podzbioru, regresja grzbietowa i lasso

Estimator	Formula
Best subset (size M)	$\hat{\beta}_j \cdot I(\lvert\hat{\beta}_j\rvert \geq \lvert\hat{\beta}_{(M)}\rvert)$
Ridge	$\hat{\beta}_j/(1+\lambda)$
Lasso	$\text{sign}(\hat{\beta}_j)(\lvert\hat{\beta}_j\rvert - \lambda)_+$

Lasso vs Ridge regression