Lasso coefficients are sparse

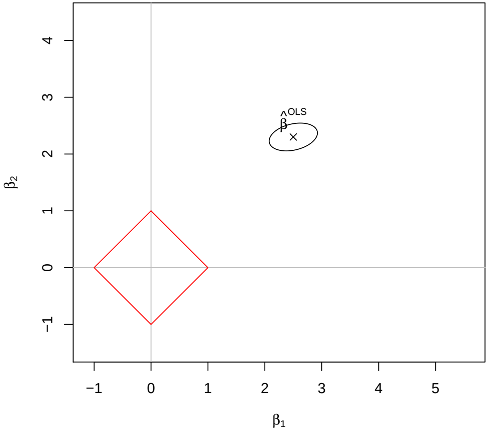


Figure: Contours of $\|\mathbf{Y} - \mathbf{X}\beta\|_2^2$ are ellipses centred at $\hat{\beta}^{OLS}$.

Lasso coefficients are sparse

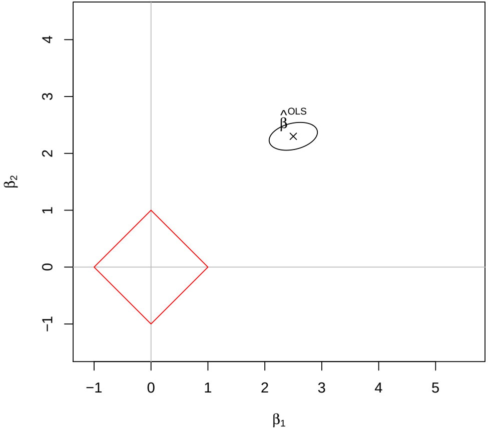
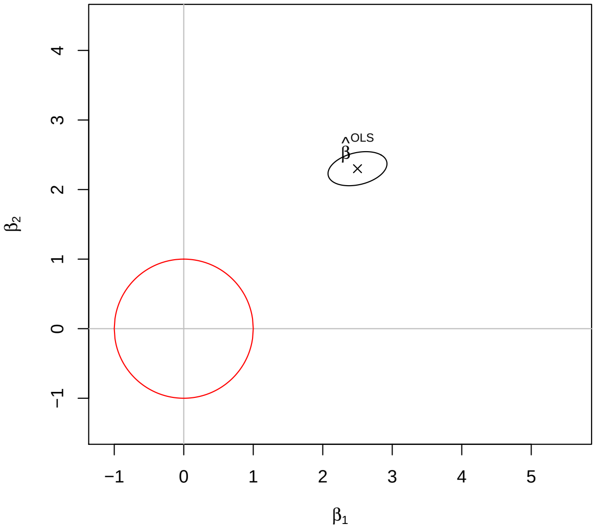


Figure: Contours of $\|\mathbf{Y} - \mathbf{X}\beta\|_2^2$ are ellipses centred at $\hat{\beta}^{OLS}$.

Ridge regression coefficients are always non-zero



Benefits of sparse coefficients

- Typically a sparse model fits well for high-dimensional data.
- Sparse models can be easier to interpret.
- In order to predict the response for a new observation, we only need measurements of a few covariates.
- Inner product $\mathbf{x}^T\hat{\beta}$ for new data point $\mathbf{x} \in \mathbb{R}^p$ fast to compute.

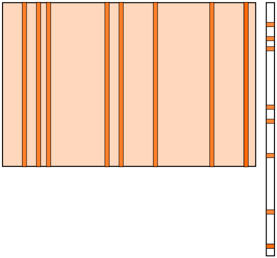
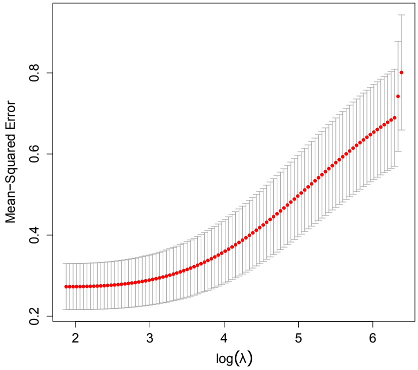


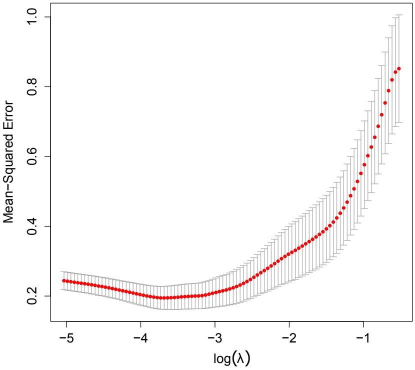
Figure: Schematic of signal $\mathbf{X}\beta^0$

Ridge regression vs the Lasso

Gene expression data, $n = 71$ observations of $p = 4088$ predictors. Response is riboflavin production by *Bacillus subtilis*.



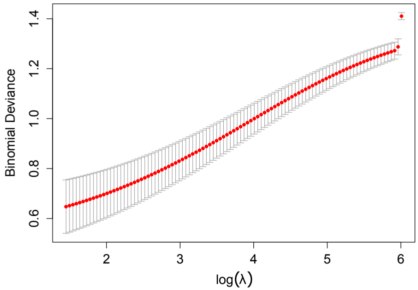
(a) Ridge regression



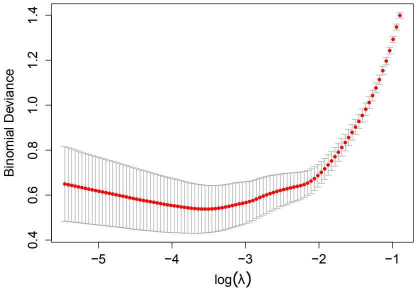
(b) Lasso regression (Tibshirani, 1996)

Ridge regression vs the Lasso

Prostate cancer gene expression data. 52 tumour samples, 50 normal samples ($n = 102$) with $p = 6033$ predictors.



(a) Ridge regression



(b) Lasso regression

ℓ_q balls

Consider penalty functions $\propto \|\boldsymbol{\beta}\|_q = \left(\sum_{k=1}^{p} \beta_k^q\right)^{1/q}$ and $p = 2$.