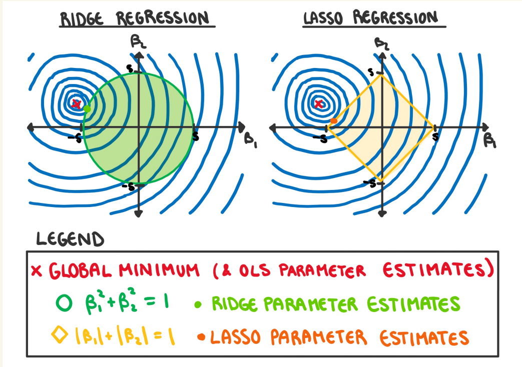
③ Shrinkage Methods (all drawings for a 2 variable model)

LASSO minimises

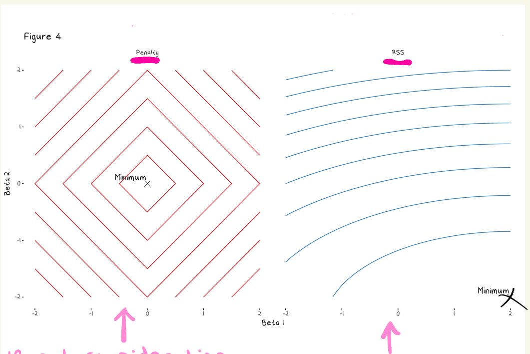
: RSS + $\lambda \sum_{j=1}^{p} |\beta_j|$

Ridge minimises

: RSS + $\lambda \sum_{j=1}^{p} \beta_j^2$

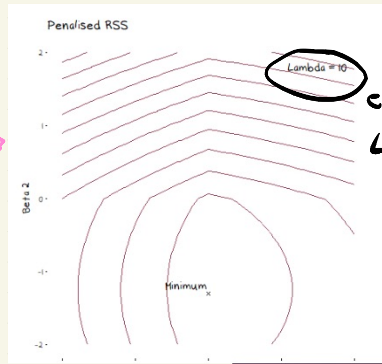


As λ increases we "shrink" the coefficients to 0. Example from the blog ↓



If only consideration, the coefficients would all be 0

If only consideration the coefficients are the OLS estimates.



e.g of Lambda

Penalised RSS is somewhere in between.

Think of λ as the weighting we give the penalty.

- This plot shows how the coefficients grow as we relax λ.

Variables that have large coefficients for all values of λ are usually more important.