Variable selection

Mark Andrews
Psychology Department, Nottingham Trent University

mark.andrews@ntu.ac.uk

Ridge regression

- Ridge regression is a method to reduce variance in estimators of regression coefficients.
- It penalizes large coefficients and shrinks the towards zero.
- In linear regression, it estimates the coefficients by minimizing the penalized sum of squared residuals:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{k=0}^{K} \beta_k^2,$$

where

$$\hat{y}_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki},$$

and λ is a regularization parameter.

Lasso

- Least absolute shrinkage and selection operator (lasso) is a method similar to ridge regression, but uses a penalty based on the sum of the *absolute* values of coefficients:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{k=0}^{K} |\beta_k|.$$

Elastic net

- Elastic net is a method similar to ridge regression and lasso.
- It uses as weighted average of the two penalty methods:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\left(\alpha\sum_{k=0}^{K}|\beta_k| + (1-\alpha)\sum_{k=0}^{K}\beta_k^2\right).$$

- The values of α ranges from 0 to 1.
- When $\alpha = 1$, this is pure lasso regression.
- When $\alpha = 0$, this is pure ridge regression.

References