H$_2$O.ai

Regularized Regression

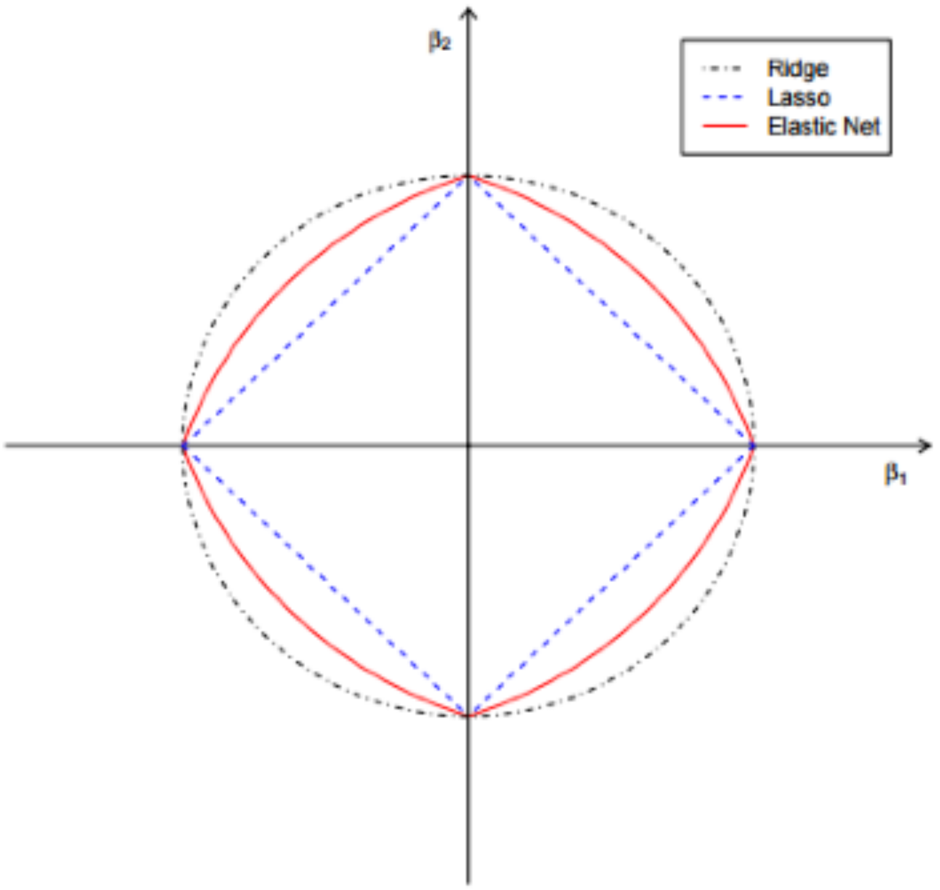
- Want to build a parsimonious but interpretable model
- Shrink the number of predictors and/or size by imposing penalties on the estimated coefficients (L1, L2)
    - LASSO:
        - picks one correlated variable, others discarded. Sparse.
    - Ridge:
        - correlated variables coefficients are pushed to the same value
    - Elastic Net:
        - sparse solution, correlated variables grouped, enter/ leave the model together

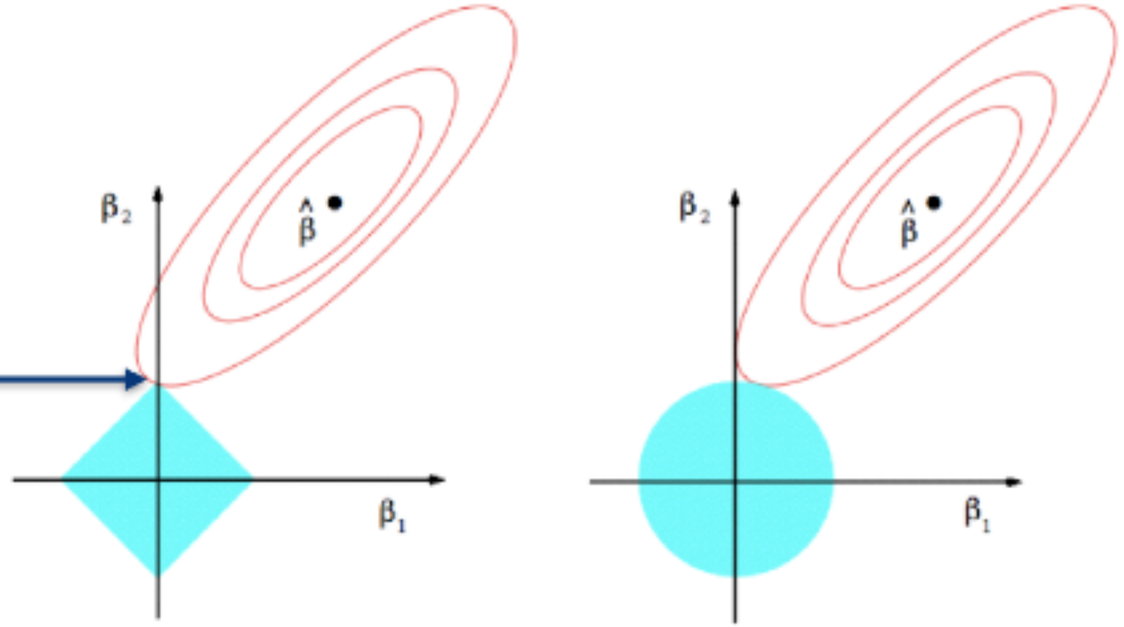
$$\min_{\beta_0,\beta} \left\{ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2 \right\} \text{ subject to } \sum_{j=1}^{p} |\beta_j| \leq t.$$

$$\text{minimize} \sum_{i=1}^{n} (y_i - \boldsymbol{\beta}^\top \mathbf{z}_i)^2 \text{ s.t. } \sum_{j=1}^{p} \beta_j^2 \le t$$

$$\arg\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1$$

Constraints for L1, L2 and L1&L2





LASSO Regression vs. Ridge Regression

Regularized Regression

- Want to build a parsimonious but interpretable model
- Shrink the number of predictors and/or size by imposing penalties on the estimated coefficients (L1, L2)
  - LASSO: $\min_{\beta_0,\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2\right\}$ subject to $\sum_{j=1}^{p}|\beta_j|\leq t.$
    - picks one correlated variable, others discarded. Sparse.
  - Ridge: $\text{minimize}\sum_{i=1}^{n}(y_i-\boldsymbol{\beta}^\top\mathbf{z}_i)^2$ s.t. $\sum_{j=1}^{p}\beta_j^2\leq t$
    - correlated variables coefficients are pushed to the same value
  - Elastic Net: $\arg\min_{\beta}\|\mathbf{y}-\mathbf{X}\beta\|^2+\lambda_2\|\beta\|^2+\lambda_1\|\beta\|_1$
    - sparse solution, correlated variables grouped, enter/ leave the model together