Regularized Regression

Shrinkage Methods

Shrinkage Methods

- Instead of Least Squares, a model is fitted using some p predictors with a technique that shrinks the coefficient estimates towards zero
- Shrinking the coefficient estimates can reduce their variance
- Ridge Regression, Lasso Regression and also Elastic Net Regression are the shrinkage methods which will be covered in this topic

Sane's STATS Academy of Statistics

ℓ_1 & ℓ_2 Forms

- Consider the elements $\beta_1, \beta_2 \dots \beta_p$.
- The ℓ_1 form of the elements, also denoted by $||\beta||_1$ is given by the expression $||\beta||_1 = \sum_{j=1}^{p} |\beta_j|$
- The ℓ_2 form of the elements, also denoted by $||\beta||_2$ is given by the expression $||\beta||_2 = \sqrt{\sum_{j=1}^{p} \beta_j^2}$

Least Squares (Quick Recap)

- In Least Squares Method, the estimates of $\beta_0, \beta_1, \beta_2 \ldots \beta_p$ are calculated by minimizing the following expression

$$Residual\ SS = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{i=1}^{p} \beta_i x_{ij})^2$$

Where

- y_i : Response value for ith observation
- x_{ij} : Value of jth observation in ith predictor

Ridge Regression

- For Ridge Regression, regression coefficients $\beta_0, \beta_1, \beta_2 \dots \beta_p$ are calculated by minimizing the following expression

  $$Residual\ SS + \lambda \sum_{j=1}^{p} \beta_j^2$$

  i.e. $\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{i=1}^{p} \beta_i x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2$

  where $\lambda \geq 0$ is a tuning parameter
- The term $\sum_{j=1}^{p} \beta_j^2$ is called shrinkage penalty. It is small if $\beta_1, \beta_2 \dots \beta_p$ are close to zero.
- For selecting optimal value for λ, cross-validation can be used

Lasso Regression

- Ridge Regression selects all the predictors in the final model. This is a disadvantage of it.
- The Lasso overcomes this disadvantage. It finds the coefficients estimates of $\beta_0, \beta_1, \beta_2 \ldots \beta_p$ by minimizing the quantity,

$$Residual\ SS + \lambda \sum_{j=1}^{p} |\beta_j|$$

i.e. $\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{i=1}^{p} \beta_i x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j|$

- The ℓ_1 penalty of Lasso has the effect of forcing some of the coefficient estimates to zero when tuning parameter λ is sufficiently large.

Variable Selection

- It can be proved that the regularized regression coefficient estimates solve the optimization problems namely,
- Minimizing $\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{i=1}^{p} \beta_i x_{ij})^2$

  subject to $\sum_{j=1}^{p}|\beta_j| \leq s$ for some finite value s (Lasso)
- Minimizing $\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{i=1}^{p} \beta_i x_{ij})^2$

  subject to$\sum_{j=1}^{p} \beta_j^2 \leq s$ for some finite value s (Ridge)

Sane's STATS Academy of Statistics

Elastic Net Regression

- Elastic Net Regression is a combination of Ridge and Lasso Regression Methods
- Elastic Net minimises:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(y - \hat{y})^2 + r\lambda \sum_{j=1}^{n}|w_j| + \frac{1-r}{2}\lambda \sum_{j=1}^{n} w_j^2$$

Sane's STATS Academy of Statistics