Regularization

Regularization

- Regularization seeks to solve a few common model issues by:
    - Minimizing model complexity
    - Penalizing the loss function
    - Reducing model overfitting (add more bias to reduce model variance)

Regularization

- In general, we can think of regularization as a way to reduce model overfitting and variance.
    - Requires some additional bias
    - Requires a search for optimal penalty hyperparameter.

Regularization

- Three main types of Regularization:
    - L1 Regularization
        - LASSO Regression
    - L2 Regularization
        - Ridge Regression
    - Combining L1 and L2
        - Elastic Net

Regularization

- L1 regularization adds a penalty equal to the **absolute value** of the magnitude of coefficients.
  - Limits the size of the coefficients.
  - Can yield sparse models where some coefficients can become zero.

Regularization

- L1 regularization adds a penalty equal to the **absolute value** of the magnitude of coefficients.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \text{RSS} + \lambda\sum_{j=1}^{p}|\beta_j|$$

Regularization

- L2 regularization adds a penalty equal to the **square** of the magnitude of coefficients.
    - All coefficients are shrunk by the same factor.
    - Does not necessarily eliminate coefficients.

Regularization

- L2 regularization adds a penalty equal to the **square** of the magnitude of coefficients.

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = \text{RSS} + \boxed{\lambda \sum_{j=1}^{p} \beta_j^2}$$

PIERIAN DATA

Regularization

- Elastic net combines L1 and L2 with the addition of an alpha parameter deciding the ratio between them:

$$\frac{\sum_{i=1}^{n}(y_i - x_i^J \hat{\beta})^2}{2n} + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^{m} \hat{\beta}_j^2 + \alpha \sum_{j=1}^{m} |\hat{\beta}_j| \right)$$

Regularization

- These regularization methods do have a cost:
    - Introduce an additional hyperparameter that needs to be tuned.
    - A multiplier to the penalty to decide the "strength" of the penalty.