Regularization
Feature scaling
Cross Validation

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

- 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

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

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

λ **: är Lambda koefficient**

Regularization – L2

- 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

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

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

Regularization – Elastic net

- 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 – Elastic net

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

$\alpha = 0$

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

Regularization – Elastic net

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

$\alpha = 1$

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

Regularization – Elastic net

- 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.

Regularization – Elastic net

- Later on, we will actually cover L2 regularization (Ridge Regression) first, due to the intuition behind the squared term being easier to understand.