Applied Machine Learning

Regression Models - Lasso & Elastic Net

Regression Models - Lasso & Elastic Net

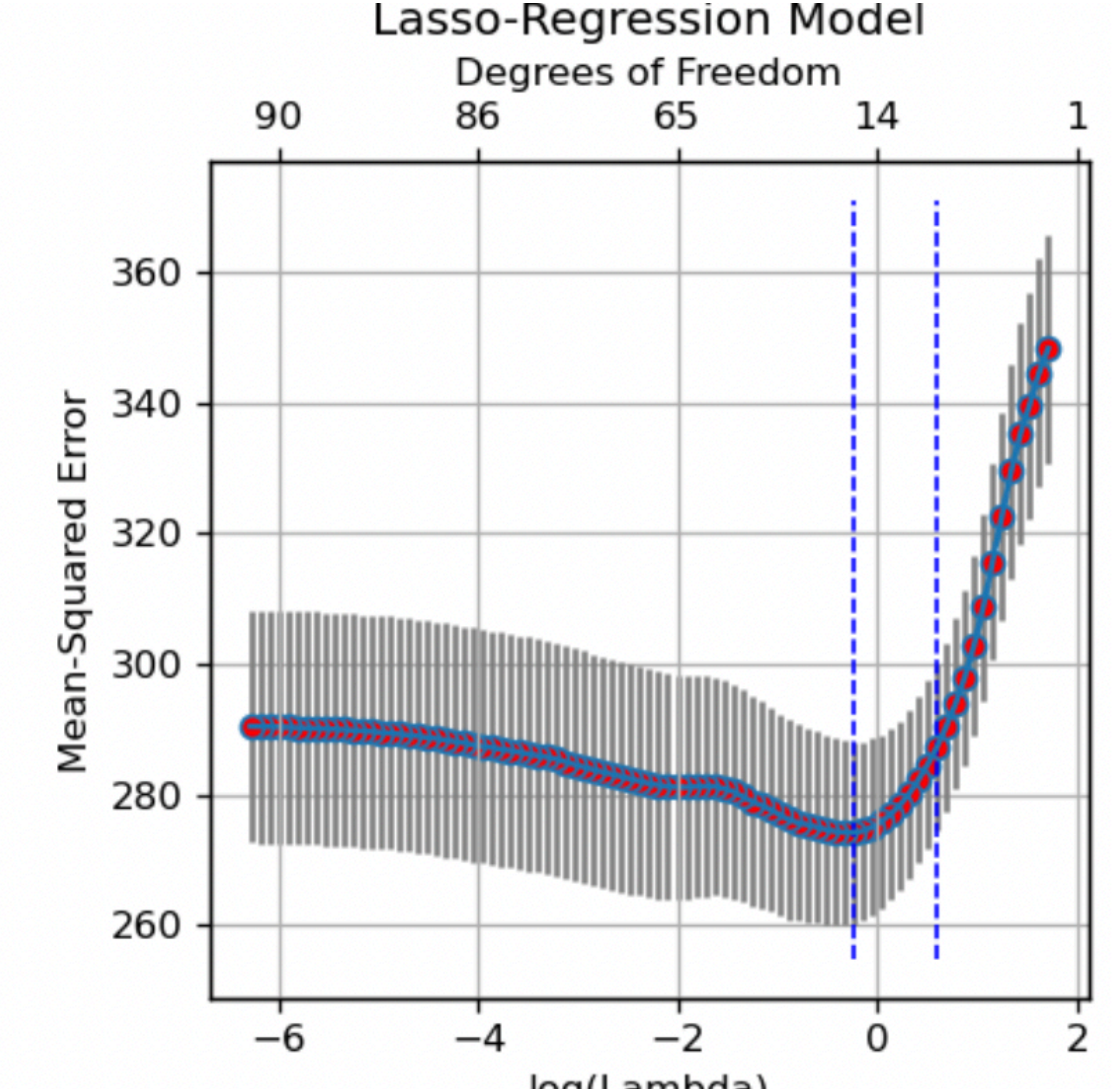
- Ridge Regularization
- Lasso Regularization
- Elastic Net

Ridge Regularization

- Minimization goal includes regularization term
  - $\frac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^\top\beta$
  - $\lambda \geq 0$ penalizes large values of β
- L_2 norm: $\|\beta\|_2 = \beta^\top\beta$
- $e(\beta_k) = (a + \lambda)\beta_k^2 - 2b(\beta_{-k})\beta_k + c(\beta_{-k}) \qquad \Rightarrow \beta_k = \dfrac{b(\beta_{-k})}{a + \lambda}$

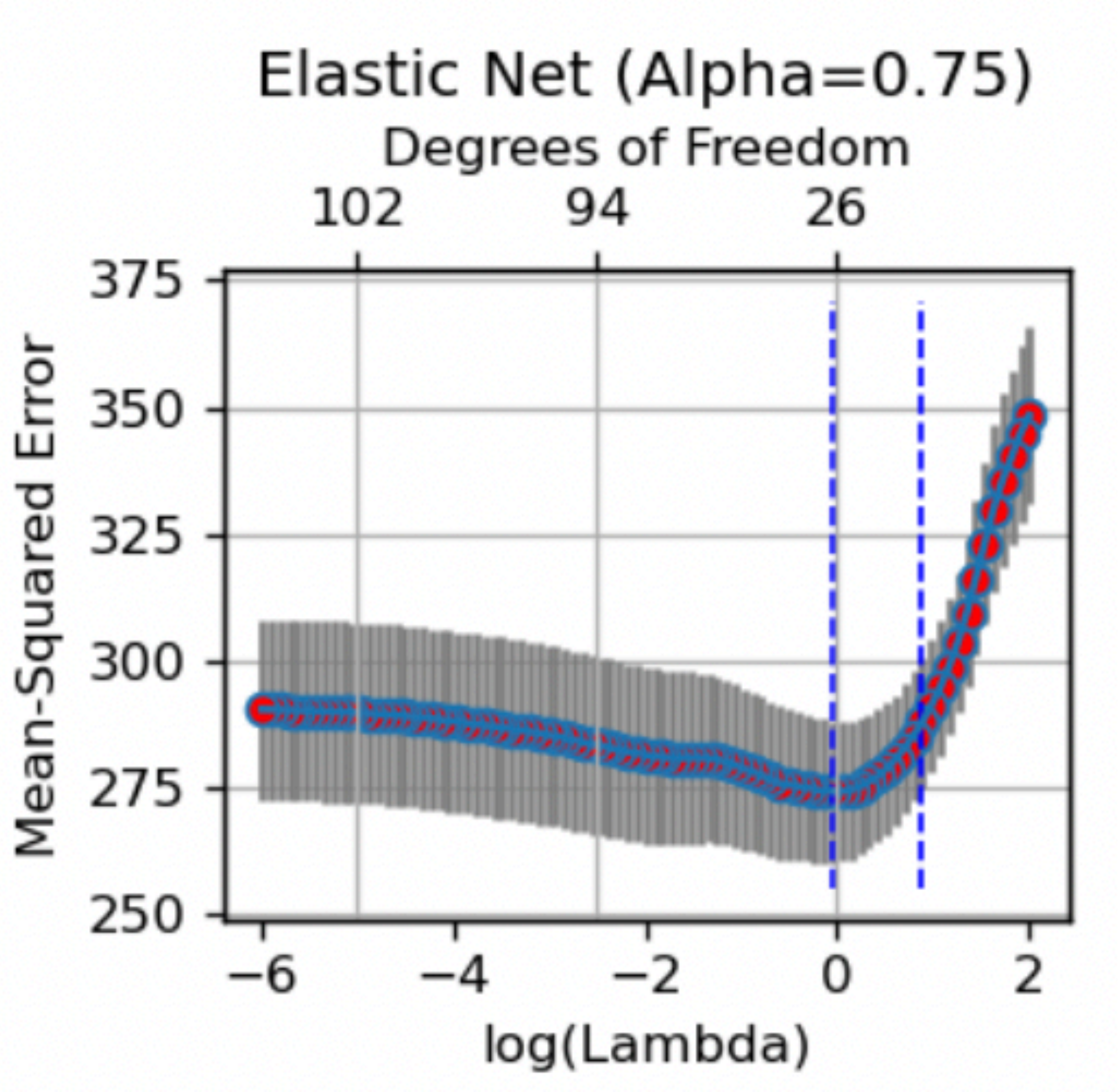
Lasso Regularization

- L_1 norm: $\|\beta\|_1 = \sum_k |\beta_k|$
- Minimization Goal:
  $$\frac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda\|\beta\|_1$$
- statistic libraries
  - stochastic descent is not a good option
  - solve regularization path for all values of $\lambda > 0$
- Explanatory variables with corresponding coefficient = 0 can be removed



Elastic Net

- Correlated explanatory variables
  - Lasso regularization may include only one
  - May result in worse predictions
- Elastic Net
  - weighted on L_1 and L_2 norms
  - minimization goal:
    $$\frac{1}{N}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) + \lambda \left(\frac{1-\alpha}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 \right)$$
  - $0 \leq \alpha \leq 1$
    - $\alpha = 1$: Lasso. $\alpha = 0$: Ridge



Regression Models - Lasso & Elastic Net

- Ridge Regularization
- Lasso Regularization
- Elastic Net

Applied Machine Learning

Regression Models - Lasso & Elastic Net