Regularization

Regularization: Lasso, Ridge, Elastic Net

Agenda:
- Regularization
- Types of Regularization
- Cross validation for GLM
- Hyper parameter tuning

Loss → $\min_{\beta_0,\beta} \left[\frac{1}{2N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T\beta)^2 + \frac{\lambda}{2} \sum_{j=1}^{p} \beta_j^2 \right]$

Ridge Regression (L2)

$\min \left[\text{sq loss} + \lambda \cdot \text{sq}\beta \right]$

overfitting

$\min_{\beta_0,\beta} \left[\text{sq loss} + \lambda \sum_{j=1}^{p} \beta_j^2 \right]$

$\beta_1^2, \beta_2^2, \beta_3^2$

Sandy Bridge

$\lambda \downarrow = 0$

$[w_1, w_2, w_3, w_4, w_5]$

$[w_1, w_2, w_3, w_4, w_5]$

Case I: $\lambda = 0$ → overfit

$\min_{\beta_0,\beta} \left[\text{sq loss} + \lambda \sum_{j=1}^{p} \beta_j^2 \right]$

not constraint

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$

$\beta_1 = 1.2$
$\beta_2 = 1.0$
$\beta_3 = 0.8$
$\beta_4 = 0.6$
$\beta_5 = 0.4$

Case II: $\lambda = 10^4$

$\min_{\beta_0,\beta} \left[\text{sq loss} + \lambda \sum_{j=1}^{p} \beta_j^2 \right]$

w_1, w_2, w_3

$\lambda \uparrow$

$\text{sum}(\beta_1^2 + \beta_2^2 + \beta_3^2) \to 0$

$\beta_1, \beta_2, \beta_3 \to 0$

$\beta_1 = 0$

$\beta_2 = 0$

$\beta_3 = 0$

Ridge : $\min_{\beta_0,\beta} \left[\frac{1}{2N} \text{SS loss} + \lambda \sum_{j=1}^{p} \beta_j^2 \right]$

$\frac{\partial}{\partial \beta} |\beta| \; \text{L1-norm}$

at origin $\frac{\partial}{\partial \beta} |\beta|$

L2-Ridge

$\frac{\partial}{\partial \beta} \beta^2$

L2-Ridge

$\frac{\partial}{\partial \beta} \beta^2$

$\beta_1 = \beta_2 = \beta_3 = 0$

$\beta_1 = 0$

$\beta_2 = 0$

$\