Regularization

Andrew Ng

Goals

- regularization
- ridge regression
- LASSO
- Elastic Net (Zou & Hastie, 2005)

Regularization

fit $\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4, \hat{y}_5$

$$\text{minimize} \quad \text{loss} + \text{regular}\left[\sum_{i=1}^{5}(\hat{y}_i - y_i)^2\right] + \left[\sum_{i=1}^{5}|\hat{y}_i - \hat{y}_{i+1}|^2 + |\hat{y}_1 - \hat{y}_5|^2\right]$$

$\underbrace{\qquad}_{\text{fit}}$ $\underbrace{\qquad}_{\text{Regularization}}$

$\text{minimum} \Rightarrow \text{value} = 0$

$\text{Regularization} \Rightarrow (\geq 0)$

$$\text{loss} = \text{minimize}\left[\sum_{i=1}^{5}(\hat{y}_i - y_i)^2\right]$$

Underfitting

$$\text{minimize} \quad \left[\underbrace{\text{loss}}_{\text{fit}} + \underbrace{\lambda \sum_{j=1}^{5} w_j^2}_{\text{Regularization}}\right]$$

$\lambda \geq 0$

λ = tuning parameter

Fit 2: $\lambda = 0$ → Overfit

Linearly weight is unrestricted. → not Overfitting

$$\text{loss} \left[\sum_{i=1}^{5}(\hat{y}_i - y_i)^2 + \lambda \sum_{j=1}^{5} w_j^2\right]$$

Overfit.

$w_1, w_2, w_3, \ldots, w_5$ = any complex number.

$w_1, w_2, w_3, \ldots, w_5 \neq 0$

$w_1 = 1.2$
$w_2 = 0.8$
$w_3 = 1.5$
$w_4 = 0.9$
$w_5 = 1.1$

Fit 3: $\lambda = \infty$

$\boxed{w_1 = 0}$ because → underfit

$$\text{minimize} \quad \left[\underbrace{\text{loss}}_{\text{fit}} + \underbrace{\lambda \sum_{j=1}^{5} w_j^2}_{\text{Regularization}}\right]$$

$\text{loss} = (1.5)^2 + (1.5)^2 + (1.5)^2 + (1.5)^2 + (1.5)^2$

$\text{loss} = 11.25$

$w_1 = w_2 = w_3 = w_4 = w_5 = 0$

$\boxed{\hat{y}_1 = \text{loss}}$

$w_1 = 5$

Differentiation

$$\text{loss} = \sum_{i=1}^{5}(\hat{y}_i - y_i)^2 = \sum_{i=1}^{5} w_i^2$$

$$\frac{\partial \text{loss}}{\partial w_1} = \frac{\partial}{\partial w_1}\sum_{i=1}^{5}(\hat{y}_i - y_i)^2 = \sum_{i=1}^{5} w_i^2$$

$$\frac{\partial \text{loss}}{\partial w_1} = \frac{\partial}{\partial w_1}\left[\sum_{i=1}^{5}(\hat{y}_i - y_i)^2 + \boxed{\lambda \sum_{j=1}^{5} w_j^2}\right]$$

Ridge Regression

$$\text{minimize} \quad \left[\underbrace{\sum_{i=1}^{5}(\hat{y}_i - y_i)^2}_{\text{fit}} + \underbrace{\lambda \sum_{j=1}^{5} w_j^2}_{\text{Regularization}}\right]$$

$\boxed{\sum_{i=1}^{5}(\hat{y}_i - y_i)^2}$ → fit

$\boxed{\lambda \sum_{j=1}^{5} w_j^2}$ → Regularization

$w_1 = w_2 = w_3 = w_4 = w_5 = 0$

$\boxed{w_1^2 = 0}$ ✓

$\boxed{\lambda \sum_{j=1}^{5} w_j^2}$ → Ridge Regularization

$w_1 = 0$ if $\lambda = \infty$

$\boxed{\text{LS-Ridge}}$

Ridge Regression

$$\text{minimize} \quad \text{MSE} + \lambda \sum_{j=1}^{5} |w_j|$$

λ = tuning parameter

$$\frac{\partial}{\partial w_1}|w_1|\Big|_{w_1=0} = \begin{cases} +1 & w_1 > 0 \\ -1 & w_1 < 0 \\ \boxed{0} & w_1 = 0 \end{cases}$$

$w_1 = 0$

$w_1 = 0$ ✓ $w_1 = 0$

$\boxed{w_1 = 0}$ → tuning is valid

$\frac{\partial}{\partial w_1}$ → numerically $|w_1|$ weight

not differentiable

$w_1 = 0$ ✓

$w_1 = 0$

Max tolerating degree of $L1$

$\boxed{L1}$ → minimize to a value but not $\boxed{0}$ (or zero).

$\boxed{L1}$ → all the two weight factor weight will be close to 0. but not $= 0$

$\boxed{\text{Not two weight factor}}$
$\boxed{\text{weight} \approx 0}$

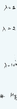
$\boxed{\text{Close to zero}}$

$\boxed{\text{Not zero}}$

$\boxed{L1}$ → all the two weight factor weight will be close to 0. but not $= 0$

$L1 = 0.1$

$$\text{loss} = \sum_{i=1}^{5}|w_i|$$



$L1 = 0.2$

$$\text{loss} = \sum_{i=1}^{5}|w_i|$$

$\boxed{\text{Strategy is reducing does not}}$
$w_2 = 0$

Relationship (L1) → Lasso Strategy MSE + L1

$\boxed{\text{Relation L1 → Lasso}}$

Strategy (L1) → $\boxed{\text{Lasso}}$ → $\boxed{\text{Relation}}$ → $\boxed{\text{Lasso}}$

$