Regularization of Linear Models

Zengchang Qin (PhD)
zengchang.qin@gmail.com

Regularization of Linear Models

1) Ridge Regression

Shrink (Penalize) the magnitude of coefficients.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \|Y - \theta^T x\|_2^2 + \lambda \|\theta\|_2^2$$

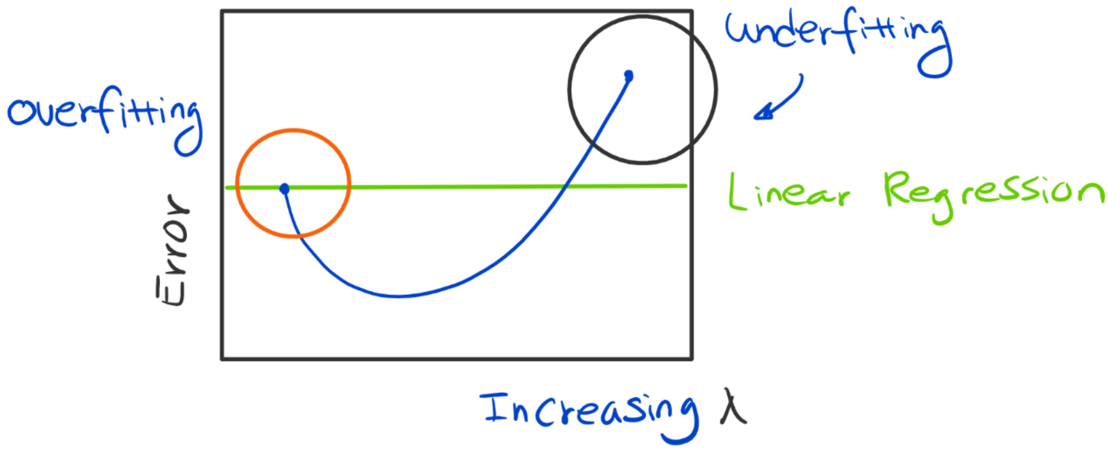
$$= \underset{\theta}{\operatorname{argmin}} \underbrace{\sum_{i=1}^{N} (y^{(i)} - \theta^T x^{(i)})^2}_{\text{Loss}} + \lambda \underbrace{\sum_{j=1}^{n} \theta_j^2}_{\text{Penalty}}$$

where λ is the parameter of penalty weight.

Not only minimizing the squared error, but also the size of the coefficients!

$\lambda = 0$, we only minimize the loss $\rightarrow$ overfitting

$\lambda = \infty$, $\theta \rightarrow 0$, minimize the penalty $\rightarrow$ underfitting

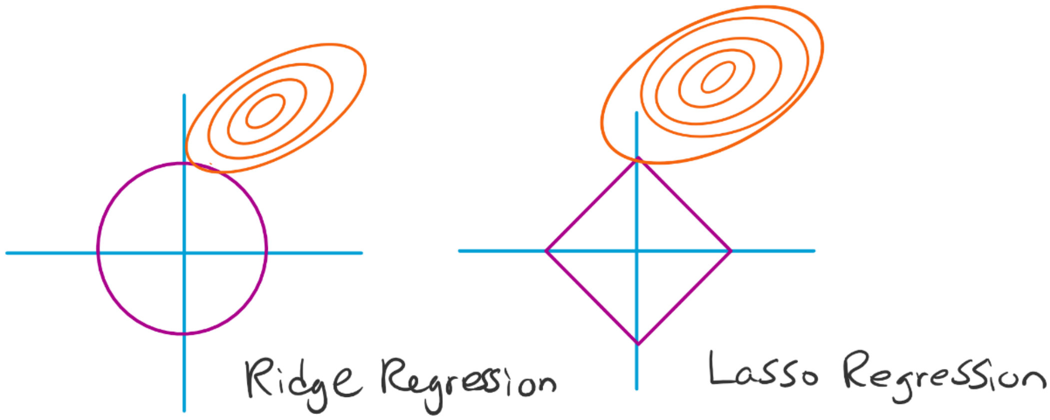


2) Lasso Regression

Lasso coefficients are defined as:

$$\hat{\theta} = \operatorname{argmin} \|Y - \theta^T X\|_2^2 + \lambda \|\theta\|_1$$

$$= \operatorname{argmin} \underbrace{\sum_{i=1}^{N} (y^{(i)} - \theta^T x^{(i)})^2}_{\text{Loss}} + \underbrace{\lambda \sum_{j=1}^{n} |\theta_j|}_{\text{Penalty}}$$



3) Solution for Ridge Regression

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - \theta^T x^{(i)})(-x_j^{(i)}) + 2\lambda \theta_j$$

$$= \frac{1}{N} \sum_{i=1}^{N} (\theta^T x^{(i)} - y^{(i)}) x_j^{(i)} + 2\lambda \theta_j$$