

# Regularization:

Good for avoiding  $\left\{ \begin{array}{l} \text{Data overfitting} \\ \text{multi-collinearity} \end{array} \right.$

$L_s$ :

$$\begin{aligned} S(\underline{\theta}) &= \sum_{i=1}^n (y_i - \underline{x}_i^T \underline{\theta})^2 \\ &= (\underline{y} - X \underline{\theta})^T (\underline{y} - X \underline{\theta}) \end{aligned}$$

Modified cost function:

$$\tilde{S}(\underline{\theta}) = \underbrace{S(\underline{\theta})}_{L_s \text{ cost}} + \underbrace{\lambda \cdot r(\underline{\theta})}_{\text{regularization term}}$$

Positive  
Regularizer

$$r(\underline{\theta}) = \begin{cases} \|\underline{\theta}\|_2^2 & \rightarrow \text{Ridge regression} \\ \|\underline{\theta}\|_1 & \rightarrow \text{Lasso regression} \\ L_p\text{-norm of } \underline{\theta} & \rightarrow \text{general form} \end{cases}$$

## Example 1: Ridge regression

$$\hat{S}(\underline{\theta}) = (\underline{y} - X\underline{\theta})^T (\underline{y} - X\underline{\theta}) + \frac{\lambda}{2} \|\underline{\theta}\|_2^2$$

Note that:  $\|\underline{\theta}\|_2^2 = \sum_i \theta_i^2$ .

$$\hat{\underline{\theta}}_{(R)} = \arg \min_{\underline{\theta}} \hat{S}(\underline{\theta})$$

"Ridge"  
=

Solving  $\nabla_{\underline{\theta}} \hat{S}(\underline{\theta}) = \underline{0}$ , yields

$$\Rightarrow -X^T(\underline{y} - X\underline{\theta}) + \lambda \cdot \underline{\theta} = \underline{0}$$

$$\Rightarrow (X^T X + \lambda I_p) \underline{\theta} = X^T \underline{y}$$

$$\Rightarrow \hat{\underline{\theta}}_{(R)} = (X^T X + \lambda I_p)^{-1} X^T \underline{y}$$

When  $\lambda = 0$ ,  $\hat{\theta}_R = \hat{\theta}_{LS} = (X^T X)^{-1} X^T y$

---

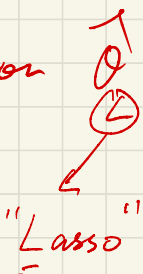
Example 2: Lasso regression

The cost function is:

$$\tilde{S}(\underline{\theta}) = S(\underline{\theta}) + \lambda \cdot |\underline{\theta}|,$$

We often need to resort to numerical algorithms.

The solution  $\hat{\theta}$  is sparse, which is favorable in many applications!



"Lasso"