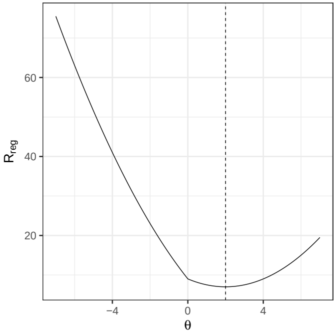
Introduction to Machine Learning

Regularization
Soft-thresholding and lasso (Deep-Dive)



Learning goals

- Understand the relationship between soft-thresholding and L1 regularization

SOFT-THRESHOLDING AND L1 REGULARIZATION

In the lecture, we wanted to solve

$$\min_{\boldsymbol{\theta}} \tilde{\mathcal{R}}_{\text{reg}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \mathcal{R}_{\text{emp}}(\hat{\theta}) + \sum_j \left[\frac{1}{2} H_{j,j}(\theta_j - \hat{\theta}_j)^2\right] + \sum_j \lambda |\theta_j|$$

with $H_{j,j} \geq 0, \lambda > 0$. Note that we can separate the dimensions, i.e.,

$$\tilde{\mathcal{R}}_{\text{reg}}(\boldsymbol{\theta}) = \sum_j z_j(\theta_j) \text{ with } z_j(\theta_j) = \frac{1}{2} H_{j,j}(\theta_j - \hat{\theta}_j)^2 + \lambda |\theta_j|.$$

Hence, we can minimize each z_j separately to find the global minimum.

If $H_{j,j} = 0$, then z_j is clearly minimized by $\hat{\theta}_{\text{lasso},j} = 0$. Otherwise, z_j is strictly convex since $\frac{1}{2} H_{j,j}(\theta_j - \hat{\theta}_j)^2$ is strictly convex and the sum of a strictly convex function and a convex function is strictly convex.