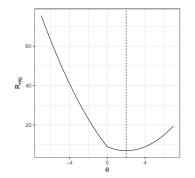
Introduction to Machine Learning

Regularization Soft-thresholding and lasso (Deep-Dive)





Learning goals

 Understand the relationship between soft-thresholding and L1 regularization

In the lecture, we wanted to solve

$$\min_{m{ heta}} ilde{\mathcal{R}}_{\mathsf{reg}}(m{ heta}) = \min_{m{ heta}} \mathcal{R}_{\mathsf{emp}}(\hat{m{ heta}}) + \sum_{j} \left[rac{1}{2} H_{j,j} (heta_j - \hat{ heta}_j)^2
ight] + \sum_{j} \lambda | heta_j|$$

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

$$ilde{\mathcal{R}}_{\mathsf{reg}}(oldsymbol{ heta}) = \sum_{j} z_{j}(heta_{j}) \; \mathsf{with} \; z_{j}(heta_{j}) = rac{1}{2} \mathcal{H}_{j,j}(heta_{j} - \hat{ heta}_{j})^{2} + \lambda | heta_{j}|.$$

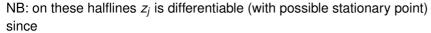
Hence, we can minimize each z_i separately to find the global minimum.

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.



For strictly convex functions, there exists only one unique minimum and for convex functions a stationary point (if it exists) is a minimum.

We now separately investigate z_j for $\theta_j > 0$ and $\theta_j < 0$.



• for
$$\theta_j > 0$$
 : $\frac{d}{d\theta_j} |\theta_j| = \frac{d}{d\theta_j} \theta_j = 1$,

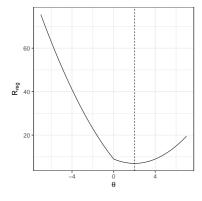
• for
$$\theta_j < 0$$
: $\frac{d}{d\theta_j} |\theta_j| = \frac{d}{d\theta_j} (-\theta_j) = -1$.



/ 2

/ 3

1)
$$\theta_j > 0$$
:



$$\frac{d}{d\theta_{j}}z_{j}(\theta_{j}) = H_{j,j}\theta_{j} - H_{j,j}\hat{\theta}_{j} + \lambda \stackrel{!}{=} 0$$

$$\Rightarrow \hat{\theta}_{\mathsf{lasso},j} = \hat{\theta}_{j} - \frac{\lambda}{H_{i,j}} > 0$$

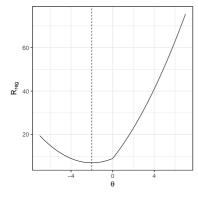
This minimum is only valid if $\hat{\theta}_{\mathrm{lasso},j} > 0$ and so it must hold that

$$\hat{\theta}_j > \frac{\lambda}{H_{i,j}}$$
.



/ 4

2)
$$\theta_j < 0$$
 :



$$\frac{d}{d\theta_j} z_j(\theta_j) = H_{j,j}\theta_j - H_{j,j}\hat{\theta}_j - \lambda \stackrel{!}{=} 0$$

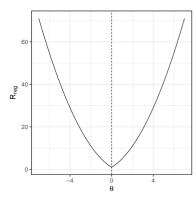
$$\Rightarrow \hat{\theta}_{\mathsf{lasso},j} = \hat{\theta}_j + \frac{\lambda}{H_{j,j}} < 0$$

This minimum is only valid if $\hat{\theta}_{{\rm lasso},j} < {\rm 0}$ and so it must hold that

$$\hat{\theta}_j < -rac{\lambda}{H_{i,i}}$$
.



/ 5



 \Rightarrow If $\hat{\theta}_j \in [-\frac{\lambda}{H_{j,j}}, \frac{\lambda}{H_{j,j}}]$ then z_j has no stationary point with

$$\hat{ heta}_{\mathsf{lasso},j} < \mathsf{0} \; \mathsf{or} \; \hat{ heta}_{\mathsf{lasso},j} > \mathsf{0}.$$

However, a unique minimum must exist since z_j is strictly convex for $H_{j,j} > 0$. This means the only possible minimizer of z_j is $\hat{\theta}_{lasso,j} = 0$.



$$\Rightarrow \hat{\theta}_{\mathsf{lasso},j} = \begin{cases} \hat{\theta}_j + \frac{\lambda}{H_{j,j}} &, \text{if } \hat{\theta}_j < -\frac{\lambda}{H_{j,j}} \text{ and } H_{j,j} > 0 \\ 0 &, \text{if } \hat{\theta}_j \in [-\frac{\lambda}{H_{j,j}}, \frac{\lambda}{H_{j,j}}] \text{ or } H_{j,j} = 0 \\ \hat{\theta}_j - \frac{\lambda}{H_{i,j}} &, \text{if } \hat{\theta}_j > \frac{\lambda}{H_{i,j}} \text{ and } H_{j,j} > 0 \end{cases}$$