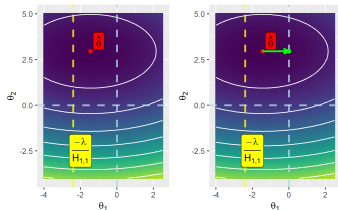# Introduction to Machine Learning

Regularization
Geometry of L1 Regularization

Learning goals

- Approximate transformation of unregularized minimizer to regularized
- Soft-Thresholding

L1-REGULARIZATION

- The L1-regularized risk of a model $f(\mathbf{x} \mid \boldsymbol{\theta})$ is

$$\mathcal{R}_{\text{reg}}(\boldsymbol{\theta}) = \mathcal{R}_{\text{emp}}(\boldsymbol{\theta}) + \sum_j \lambda |\theta_j|$$

and the (sub-)gradient is:

$$\nabla_{\boldsymbol{\theta}} \mathcal{R}_{\text{emp}}(\boldsymbol{\theta}) + \lambda \, \text{sign}(\boldsymbol{\theta})$$

- Unlike in $L2$, contribution to grad. doesn't scale with θ_j elements.
- Again: quadratic Taylor approximation of $\mathcal{R}_{\text{emp}}(\boldsymbol{\theta})$ around its minimizer $\hat{\theta}$, then regularize:

$$\tilde{\mathcal{R}}_{\text{reg}}(\boldsymbol{\theta}) = \mathcal{R}_{\text{emp}}(\hat{\theta}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\theta})^T \boldsymbol{H}(\boldsymbol{\theta} - \hat{\theta}) + \sum_j \lambda |\theta_j|$$