Exercise of Supervised Learning: Regularization Part 1

Yawei Li

yawei.li@stat.uni-muenchen.de

December 8, 2023

Exercise 1: L0 Regularization

Consider the regression learning setting, i.e., $\mathcal{Y} = \mathbb{R}$, and the feature space $\mathcal{X} = \mathbb{R}^p$. Let the hypothesis space be the linear models:

$$\mathcal{H} = \{f(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x} \mid \boldsymbol{\theta} \in \mathbb{R}^p\}.$$

Suppose your loss function of interest is the L2 loss $L(y, f(\mathbf{x})) = \frac{1}{2}(y - f(\mathbf{x}))^2$. Consider the L_0-regularized empirical risk of a model $f(\mathbf{x} \mid \boldsymbol{\theta})$:

$$\mathcal{R}_{\text{reg}}(\boldsymbol{\theta}) = \mathcal{R}_{\text{emp}}(\boldsymbol{\theta}) + \lambda ||\boldsymbol{\theta}||_0 = \frac{1}{2} \sum_{i=1}^{n} (y^{(i)} - \boldsymbol{\theta}^T \mathbf{x}^{(i)})^2 + \lambda \sum_{i=1}^{p} \boldsymbol{I}_{|\theta_i| \neq 0}.$$

Assume that $\mathbf{X}^T\mathbf{X} = \mathbf{I}$, which holds if $\mathbf{X}$ has orthonormal columns. Show that the minimizer $\hat{\theta}_{L0} = (\hat{\theta}_{L0,1}, \dots, \hat{\theta}_{L0,p})^T$ is given by

$$\hat{\theta}_{L0,i} = \hat{\theta}_i \boldsymbol{I}_{\hat{\theta}_i > \sqrt{2\lambda}}, \qquad i = 1, \dots, p,$$

where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_p)^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ is the minimizer of the unregularized empirical risk. For this purpose, using the following steps:

Exercise 1 (i)

(i) Derive that

$$\arg\min_{\boldsymbol{\theta}} \mathcal{R}_{\text{emp}}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{p} -\hat{\theta}_i\theta_i + \frac{\theta_i^2}{2} + \lambda \boldsymbol{I}_{|\theta_i| \neq 0}.$$

Solution to Exercise 1 (i)