Penalized Non-Linear Least Squares

Ilias Moysidis

Centre for Research and Technology - Hellas

1 Model Description

Let $\|\cdot\|_2$ denote the usual euclidean norm. Consider the penalized least squares problem

$$\underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \|\mathbf{z}-\mathbf{h}(\mathbf{x},\boldsymbol{\beta})\|_2^2+\lambda\|\boldsymbol{\beta}\|_2^2, \tag{1.1}$$

where $\mathbf{z},\mathbf{x},\mathbf{h}(\mathbf{x},\boldsymbol{\beta})\in\mathbb{R}^n$. The idea is to approximate $\mathbf{h}(\mathbf{x},\boldsymbol{\beta})$ with the help of the Taylor expansion theorem as

$$\mathbf{h}(\mathbf{x},\boldsymbol{\beta})\approx\mathbf{h}(\mathbf{x},\boldsymbol{\beta}^{(t)})+\mathbf{J}(\mathbf{x},\boldsymbol{\beta}^{(t)})(\boldsymbol{\beta}-\boldsymbol{\beta}^{(t)}),\quad \mathbf{J}(\mathbf{x},\boldsymbol{\beta}^{(t)})=\frac{\partial\mathbf{h}}{\partial\boldsymbol{\beta}}(\mathbf{x},\boldsymbol{\beta}^{(t)})\in\mathbb{R}^{n\times p}.$$

What we achieved here is the linearization of $\mathbf{h}$ and therefore the transformation of (1.1) to the known penalized least squares problem

$$\boldsymbol{\beta}^{(t+1)}=\underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \|\mathbf{y}-\mathbf{J}(\mathbf{x},\boldsymbol{\beta}^{(t)})\boldsymbol{\beta}\|_2^2+\lambda\|\boldsymbol{\beta}\|_2^2,\quad \mathbf{y}=\mathbf{z}-\mathbf{h}(\mathbf{x},\boldsymbol{\beta}^{(t)})+\mathbf{J}(\mathbf{x},\boldsymbol{\beta}^{(t)})\boldsymbol{\beta}^{(t)}. \tag{1.2}$$

We repeat this procedure until convergence. This algorithm is known as Gauss-Newton. The iterations of the algorithm are given by the solution of (1.2):

$$\boldsymbol{\beta}^{(t+1)}=\left[\mathbf{J}(\mathbf{x},\boldsymbol{\beta}^{(t)})^\intercal\mathbf{J}(\mathbf{x},\boldsymbol{\beta}^{(t)})+\lambda\mathbf{I}_p\right]^{-1}\mathbf{J}(\mathbf{x},\boldsymbol{\beta}^{(t)})^\intercal\mathbf{y}.$$

1.1 Example

As a proof of concept we can try the algorithm on the model function

$$h(x, \boldsymbol{\beta}) = \sum_{i=1}^{p} \cos(i\beta_i x).$$

The jacobian matrix of h is given by

$$\frac{\partial h}{\partial \beta_i}(x, \boldsymbol{\beta}) = -ix \sin(i\beta_i x).$$

We randomly generate covariates $\boldsymbol{\beta}$ from $\mathcal{N}_p(\mathbf{0}, 0.5^2\mathbf{I})$, data $\mathbf{x}$ from $\mathcal{N}_n(\mathbf{0}, \mathbf{I})$ and noise from $\mathcal{N}_n(\mathbf{0}, 0.01^2\mathbf{I})$. We choose $\lambda = 2 \cdot 10^{-9}$, $n = 20$, $p = 30$.

References