Statistica Sinica Preprint No: SS-2016-0222R2

Title	Gradient-induced Model-free Variable Selection with Composite Quantile Regression
Manuscript ID	SS-2016.0222
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0222
Complete List of Authors	Shaogao Lv Xin He and Junhui Wang
Corresponding Author	Shaogao Lv
E-mail	lvsg716@swufe.edu.cn
Notice: Accepted version subject to English editing.	

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0222

and its empirical version as

$$\mathcal{E}_{\mathcal{Z}}(\mathbf{Q},\mathbf{g}) = \frac{1}{mn(n-1)} \sum_{k=1}^{m} \sum_{i,j=1}^{n} w_{ij} L_{\tau_k}\Big(y_i - Q_{\tau_k}(\mathbf{x}_j) - \mathbf{g}_{\tau_k}(\mathbf{x}_i)^T(\mathbf{x}_i - \mathbf{x}_j)\Big).$$

The proposed method is then formulated as

$$\underset{\mathbf{Q},\mathbf{g}}{\operatorname{argmin}} \ \mathcal{E}_{\mathcal{Z}}(\mathbf{Q},\mathbf{g}) + \frac{\lambda_0}{m}\sum_{k=1}^{m} \|Q_{\tau_k}\|^2_{\mathcal{H}_K} + \lambda_1 \sum_{l=1}^{p} \pi_l \|\mathbf{g}^l\|_{\mathcal{H}_K^m}. \tag{3}$$

Here $\|\mathbf{g}^l\|_{\mathcal{H}_K^m} = \sqrt{\frac{1}{m}\sum_{k=1}^{m} \|g^l_{\tau_k}\|^2_{\mathcal{H}_K}}$ is a group Lasso penalty (Yuan and Lin (2006)) that has the effect of pushing all or none of elements in $\|\mathbf{g}^l\|_{\mathcal{H}_K^m}$ to be exactly 0, thus achieving the purpose of variable selection. The weight π_l is adaptively assigned to different $\|\mathbf{g}^l\|_{\mathcal{H}_K^m}$ to achieve better selection performance following the suggestion of Zou and Yuan (2008), the penalty term $\|Q_{\tau_k}\|^2_{\mathcal{H}_k}$ is a standard RKHS-norm penalty, and λ_0 and λ_1 are two tuning parameters.

2.3 Computing algorithm

In this section, we develop an efficient computing algorithm to solve (3) , which couples the MM algorithm and the proximal gradient descent algorithm. The algorithm proceeds in an iterative fashion. Given the current estimate $(\widetilde{\mathbf{Q}}, \widetilde{\mathbf{g}})$ and $\tilde{o}_{ij} = y_i - \widetilde{Q}_{\tau_k}(\mathbf{x}_j) - \widetilde{\mathbf{g}}_{\tau_k}(\mathbf{x}_i)^T(\mathbf{x}_i - \mathbf{x}_j)$, we first approximate the check loss $L_{\tau_k}(o_{ij})$ with a smooth loss function $L^{\epsilon}_{\tau_k}(o_{ij}) = L_{\tau_k}(o_{ij}) - \frac{\epsilon}{2}\ln(\epsilon + |o_{ij}|)$ and then majorize it with

$$\widetilde{L}^{\epsilon}_{\tau_k}(o_{ij}|\tilde{o}_{ij}) = \frac{1}{4}\left(\frac{o_{ij}^2}{\epsilon + |\tilde{o}_{ij}|} + (4\tau_k - 2)o_{ij} + c\right),$$

where c is a constant such that $\widetilde{L}^{\epsilon}_{\tau_k}(\tilde{o}_{ij}|\tilde{o}_{ij}) = L^{\epsilon}_{\tau_k}(\tilde{o}_{ij})$. The minimization step is then to solve

$$\underset{\mathbf{Q},\mathbf{g}}{\operatorname{argmin}} \ R(\mathbf{Q},\mathbf{g}) + \frac{\lambda_0}{m}\sum_{k=1}^{m} \|Q_{\tau_k}\|^2_{\mathcal{H}_K} + \Omega(\mathbf{g}), \tag{4}$$