Journal of the Korean Data & Information Science Society
2014, **25**(3), 677–684

http://dx.doi.org/10.7465/jkdi.2014.25.3.677
한국데이터정보과학회지

Support vector quantile regression ensemble with bagging[†]

Jooyong Shim[1] · Changha Hwang[2]

[1]Department of Data Science, Inje University
[2]Department of Applied Statistics, Dankook University

Received 27 April 2014, revised 12 May 2014, accepted 18 May 2014

Abstract

Support vector quantile regression (SVQR) is capable of providing more complete description of the linear and nonlinear relationships among random variables. To improve the estimation performance of SVQR we propose to use SVQR ensemble with bagging (bootstrap aggregating), in which SVQRs are trained independently using the training data sets sampled randomly via a bootstrap method. Then, they are aggregated to obtain the estimator of the quantile regression function using the penalized objective function composed of check functions. Experimental results are then presented, which illustrate the performance of SVQR ensemble with bagging.

Keywords: Bootstrap aggregating, check function, cross validation function, kernel function, support vector quantile regression.

1. Introduction

Since Koenker and Bassett (1978) introduced linear quantile regression, quantile regression has been a popular method for estimating the quantiles of a conditional distribution given input variables. Just as classical linear regression methods based on minimizing sum of squared residuals enable us to estimate a wide variety of models for conditional mean functions, quantile regression methods offer a mechanism for estimating models for the full range of conditional quantile functions, including the conditional median function. By supplementing the estimation of conditional mean functions with techniques for estimating an entire family of conditional quantile functions, quantile regression is capable of providing a better statistical analysis of the stochastic relationships among random variables. An introduction and look at current research areas of quantile regression can be found in Koenker and Hallock (2001), Yu *et al*. (2003), Koenker (2005), Hwang (2010), Lee and Shim (2010), and Shim and Hwang (2012, 2013).

[†] This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A3A2033330).

[1] Adjunct professor, Institute of Statistical Information, Department of Data Science, Inje University, Kyungnam 621-749, Korea.

[2] Corresponding author: Professor, Department of Statistics, Dankook University, Gyeonggido 448-701, Korea. E-mail: chwang@dankook.ac.kr

Support vector machine (SVM), firstly developed by Vapnik (1995), is being used as a new technique for regression and classification problems. For applications of SVM see Hwang and Shim (2012) and Shim and Hwang (2012, 2013). SVM is based on the structural risk minimization (SRM) principle, which has been shown to be superior to traditional empirical risk minimization (ERM) principle (Vapnik, 1995). SRM minimizes an upper bound on the expected risk unlike ERM minimizing the error on the training data. Support vector quantile regression (SVQR) can be obtained by applying SVM with a check function instead of an ε -insensitive loss function into the quantile regression (Hwang and Shim, 2005; Takeuchi *et al.*, 2006).

The bootstrap method is a computer based method for assigning measures of accuracy to statistical estimates, which generates a large number of bootstrap samples by repeatedly resampling the original data set in random manner to provide informations on the distribution of the statistic of interest. A good introduction can be found in Efron and Tibshirani (1993). SVM ensemble has been proposed by Kim *et al.* (2003), in which the bagging (bootstrap aggregating; Breiman, 1996) and the boosting (Schapire *et al.*, 1998) are used to train individual SVM and combine several SVMs.

In this paper we propose to use the SVQR ensemble with bagging to obtain better estimator of the quantile regression function. SVQRs are trained independently using the training data sets sampled randomly from the original training data set via the bootstrap method. They are aggregated to obtain the estimator of the quantile regression function given original training data or test data in the form of the weighted sum of each estimator of the quantile regression function obtained by using each bootstraped training data set.

The rest of this paper is organized as follows. In Section 2 the support vector quantile regression is briefly introduced. In Section 3 the support vector quantile regression ensemble with bagging is proposed. In Section 4 we perform the numerical studies through artificial and real examples. In Section 5 we give the conclusions.

2. Support vector quantile regression

Let the training data set be denoted by $(\boldsymbol{x}_{i,}y_i)_{i=1}^n$, with each input $\boldsymbol{x}_i \in R^d$ and the response $y_i \in R$, where the response variable y_i is nonlinearly related to the input vector $\boldsymbol{x}_i$. Here the feature mapping function $\phi(\cdot) : R^d \to R^{d_f}$ maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way. An inner product in feature space has an equivalent kernel in input space, $\phi(\boldsymbol{x}_i)'\phi(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ (Mercer, 1909). Several choices of the kernel $K(\cdot, \cdot)$ are possible. We consider the nonlinear regression case, in which the quantile regression function $q(\boldsymbol{x}_t)$ of the response given $\boldsymbol{x}_t$ can be regarded as a nonlinear function of input vector $\boldsymbol{x}_t$.

With a check function $\rho_\theta(\cdot)$, the estimator of the θth quantile regression function can be defined as any solution to the optimization problem,

$$\min \ell(q_\theta|\boldsymbol{x}) = \sum_{i=1}^{n} \rho_\theta(y_i - q(\boldsymbol{x}_i)) \tag{2.1}$$

where $\rho_\theta(r) = \theta r I_{(r\geq 0)} + (1-\theta) r I_{(r<0)}$.

We can express the regression problem by formulation for SVM as follows.

$$\min L = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}(\theta\xi_i + (1-\theta)\xi_i^*) \tag{2.2}$$

subject to

$$y_i - \boldsymbol{w}'\phi(\boldsymbol{x}_i) - b \le \xi_i, \boldsymbol{w}'\phi(\boldsymbol{x}_i) + b - y_i \le \xi_i^*,\ \xi_i,\ \xi_i^* \ge 0,$$

where C is a positive penalty parameter penalizing the training errors.We construct a Lagrange function as follows:

$$\begin{aligned} L =& \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}(\theta\xi_i + (1-\theta)\xi_i^*) - \sum_{i=1}^{n}\alpha_i(\xi_i - y_i + \boldsymbol{w}'\phi(\boldsymbol{x}_i) + b) \\ &- \sum_{i=1}^{n}\alpha_i^*(\xi_i^* + y_i - \boldsymbol{w}'\phi(\boldsymbol{x}_i) - b) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*). \end{aligned} \tag{2.3}$$

We notice that the positivity constraints $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \ge 0$ should be satisfied. After taking partial derivatives of equation (2.3) with regard to the primal variables $(\boldsymbol{w}, b, \xi_i, \xi_i^*)$ and plugging them into equation (2.3), we have the optimization problem below.

$$\max\ -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i \tag{2.4}$$

with constraints

$$0 \le \alpha_i \le \theta C, 0 \le \alpha_i^*(1-\theta)C \text{ and } \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0.$$

Solving the above equation with the constraints determines the optimal Lagrange multipliers, α_i, α_i^*,the estimator of the θth quantile regression function given the input vector $\boldsymbol{x}_t$ are obtained as follows:

$$\widehat{q}_\theta(\boldsymbol{x}_t) = \sum_{i=1}^{n}K(\boldsymbol{x}_t, \boldsymbol{x}_i)(\widehat{\alpha}_i - \widehat{\alpha}_i^*) + \widehat{b}. \tag{2.5}$$

Here $\widehat{b}$ is obtained via Kuhn-Tucker conditions (Kuhn and Tucker, 1951) such as,

$$\widehat{b} = \frac{1}{n_s}\sum_{i \in I_s}(y_i - K(\boldsymbol{x}_i, \boldsymbol{x})(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}*)), \tag{2.6}$$

where $\boldsymbol{x} = \{\boldsymbol{x}_i\}_{i=1}^n$, $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, \cdots, \hat{\alpha}_n)'$, $\widehat{\boldsymbol{\alpha}}^* = (\widehat{\alpha}_1^*, \cdots, \hat{\alpha}_n^*)'$ and n_s is the size of the set $I_s = \{i = 1, \cdots, n|\ 0 < \widehat{\alpha}_i < C\theta, 0 < \hat{\alpha}_i^* < C(1-\theta)\}$.

The functional structures of SVQR is characterized by the hyper-parameters, C and the kernel parameters. To select the hyper-parameters of SVQR we consider the cross validation (CV) function defined as follows:

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\rho_\theta(y_i - \widehat{q}_\theta(\boldsymbol{x}_i)^{(-i)}), \tag{2.7}$$

where λ is the set of hyper-parameters and $\widehat{q}_\theta(\boldsymbol{x}_i)^{(-i)}$ is the quantile regression function estimated without ith observation. Since for each candidates of parameters, $\widehat{q}_\theta(\boldsymbol{x}_i)^{(-i)}$ for

$i = 1, \cdots, n$, should be evaluated, selecting parameters using CV function is computationally formidable. Yuan (2006) proposed the generalized approximate cross validation (GACV) function to select the set of hyper-parameters λ for SVQR as follows:

$$GACV(\lambda) = \frac{\sum_{i=1}^{n} \rho_\theta(y_i - \hat{q}_\theta(\boldsymbol{x}_i))}{n - trace(H)}, \tag{2.8}$$

where H is the hat matrix such that $\hat{q}(\theta|\boldsymbol{x}) = H\boldsymbol{y}$ with the (i,j)th element $h_{ij} = \frac{\partial \hat{q}_\theta(\boldsymbol{x}_i)}{\partial y_j}$. From Li *et al.* (2007) we have that the trace of the hat matrix H equals to the size of set I_s used in (2.6).

3. Ensemble SVQR with bagging

We denote the jth bootstrapped training data set by $(\boldsymbol{x}_i^{(j)}, y_i^{(j)})_{i=1}^n$, which is randomly sampled with replacement from the training data set $(\boldsymbol{x}_i, y_i)_{i=1}^n$, and denote the estimator of the θth quantile regression function given $\boldsymbol{x}_i$ using the jth bootstrapped training data set by $\widehat{q}_\theta^{(j)}(\boldsymbol{x}_i)$ for $j = 1, \cdots, n_B$. Using $\{\widehat{q}_\theta^{(j)}(\boldsymbol{x}_i)\}_{j=1}^{n_B}$ we want to find the estimator of the θth quantile regression function given $\boldsymbol{x}_i$, $\widehat{q}_\theta^B(\boldsymbol{x}_i)$, which is the weighted sum of $\widehat{q}_\theta^{(j)}(\boldsymbol{x}_i)$'s as follows:

$$\widehat{q}_\theta^B(\boldsymbol{x}_i) = \sum_{j=1}^{n_B} u_j \widehat{q}_\theta^{(j)}(\boldsymbol{x}_i), \tag{3.1}$$

where u_j is weight to be estimated.

To obtain the estimate of weight $u_j, j = 1, \cdots, n_B$, we first consider the objective function as follows:

$$\ell = \sum_{i=1}^{n} \rho_\theta(y_i - \widehat{q}_\theta^B(\boldsymbol{x}_i)) = \sum_{i=1}^{n} \rho_\theta(y_i - \boldsymbol{u}'\widehat{Q}_\theta(\boldsymbol{x}_i))$$

where $\widehat{Q}_\theta(\boldsymbol{x}_i)$ is the ith column of $\widehat{Q}_\theta(\boldsymbol{x}) = \{\widehat{q}_\theta^{(j)}(\boldsymbol{x}_i)\}_{j=1}^{n_B}{}_{i=1}^{n}$.

Since $\widehat{Q}_\theta(\boldsymbol{x})$ is a $n \times n_B$ matrix and n_B is usually greater than n, the weight vector $\boldsymbol{u} = \{u_j\}_{j=1}^K$ can not be obtained by minimizing ℓ. Thus we consider the penalized objective function motivated by ridge regression (Saunders *et al.*, 1998) as follows:

$$\ell = \frac{1}{2}||\boldsymbol{u}||^2 + C_0 \sum_{i=1}^{n} \rho_\theta(y_i - \boldsymbol{u}'\widehat{Q}_\theta(\boldsymbol{x}_i)),$$

subject to

$$u_j \geq 0 \text{ for } j = 1, \cdots, n_B,$$

where $C_0 > 0$ is a penalty parameter.

We can express the above optimization problem by formulation of SVQR as follows:

$$\min \frac{1}{2}||\boldsymbol{u}||^2 + C_0 \sum_{i=1}^{n} (\theta\xi_i + (1-\theta)\xi_i^*) \tag{3.2}$$

subject to

$$y_i - \boldsymbol{u}'\widehat{Q}_\theta(\boldsymbol{x}_i) \le \xi_i \, \boldsymbol{u}'\widehat{Q}_\theta(\boldsymbol{x}_i) - y_i \le \xi_i^*, \ \xi_i^{(*)} \ge 0, \ u_j \ge 0.$$

We construct a Lagrange function as follows:

$$L = \frac{1}{2}||\boldsymbol{u}||^2 + C_0 \sum_{i=1}^n (\theta\xi_i + (1-\theta)\xi_i^*) - \sum_{i=1}^n \alpha_i(\xi_i - y_i + \boldsymbol{u}'\widehat{Q}_\theta(\boldsymbol{x}_i)) \tag{3.3}$$
$$- \sum_{i=1}^n \alpha_i^*(\xi_i^* - \boldsymbol{u}'\widehat{Q}_\theta(\boldsymbol{x}_i) + y_i) - \sum_{i=1}^n \eta_i\xi_i - \sum_{i=1}^n \eta_i^*\xi_i^* - \sum_{j=1}^{n_B} v_j u_j.$$

We notice that the non-negative constraints $\alpha_i^{(*)}, \eta_i^{(*)}, v_j \ge 0$ should be satisfied. After taking partial derivatives of the equation (3.3) with respect to the primal variables $(\boldsymbol{u}, \xi_i, \xi_i^*)$ and plugging them into the equation (3.3), we have the optimization problem below.

$$\min\ L = \frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)'\widehat{Q}_\theta(\boldsymbol{x})\widehat{Q}_\theta(\boldsymbol{x})'(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*) - \boldsymbol{v}'\widehat{Q}_\theta(\boldsymbol{x})'(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*) + \frac{1}{2}\boldsymbol{v}'\boldsymbol{v} + \boldsymbol{y}'(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*) \tag{3.4}$$

subject to

$$0 \le \alpha_i \le \theta C_0 \text{ and } 0 \le \alpha_i^* \le (1-\theta)C_0,$$

where $\boldsymbol{\alpha}^{(*)} = (\alpha_1^{(*)}, \cdots, \alpha_n^{(*)})'$, $\boldsymbol{y} = (y_1, \cdots, y_n)'$ and $\boldsymbol{v} = (v_1, \cdots, v_{n_B})'$.

Solving the above problem with the constraints determines the optimal Lagrange multipliers $\hat{\alpha}_i$ and $\hat{\alpha}_i^*$. Thus, the estimated $\boldsymbol{u}$ is obtained as

$$\widehat{\boldsymbol{u}} = \widehat{Q_\theta}(\boldsymbol{x}_i)(\widehat{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\alpha}}*) + \widehat{\boldsymbol{v}}. \tag{3.5}$$

Thus the estimator of the θth quantile regression function given $\boldsymbol{x}_t$ is obtained as follows;

$$\widehat{q}_\theta^B(\boldsymbol{x}_t) = \sum_{j=1}^{n_B} \hat{u}_j \widehat{q}_\theta^{(j)}(\boldsymbol{x}_t), \tag{3.6}$$

where $\widehat{q}_\theta^{(j)}(\boldsymbol{x}_t)$ is the estimator of the θth quantile regression function given $\boldsymbol{x}_t$ using the jth the bootstrapped training data set $(\boldsymbol{x}_i^{(j)}, y_i^{(j)})_{i=1}^n$.

The functional structures of SVQR ensemble with bagging is characterized by the hyper-parameters, C_0, C and the kernel parameters. To select the hyper-parameters of SVQR ensemble with bagging we consider the leave- one-out cross validation (CV) function as follows:

$$CV(C) = \frac{1}{n}\sum_{i=1}^n \rho_\theta(y_i - \widehat{q}_\theta^B(\boldsymbol{x}_i)^{(-i)}), \tag{3.7}$$

where $\widehat{q}_\theta^B(\boldsymbol{x}_i)^{(-i)}$ is the estimator of the θth quantile regression function given $\boldsymbol{x}_i$ obtained without ith observation $(\boldsymbol{x}_i, y_i)$. For fast computation we use k-fold cross validation (kCV) function as follows:

$$kCV(C) = \sum_{l=1}^k \frac{1}{n_l}\sum_{i \in F_l} \rho_\theta(y_i - \widehat{q}_\theta^B(\boldsymbol{x}_i)^{(-l)}), \tag{3.8}$$

where $\widehat{q}_\theta^B(\boldsymbol{x}_i)^{(-l)}$ is estimator of the θth quantile regression function give $\boldsymbol{x}_i$ obtained without data in the lth subset, F_l is the lth subset such that $\bigcup_{l=1}^k F_l = \{1, 2, \cdots, n\}$ and $F_i \cap F_j = \{\}$ for $i \ne j$, n_l is the size of F_l such that $n = \sum_{l=1}^k n_l$.

4. Numerical studies

In this section, we illustrate the performance of SVQR ensemble with bagging through the simulated examples on the nonlinear quantile regression case. In the single SVQR the optimal values of C and σ^2 are chosen by GACV function (2.8), in SVQR ensemble with bagging (SVQR ensemble) the optimal values of C_0, C and σ^2 are chosen by 5-fold CV function (3.8). We generate 100 training data sets and use 200 bootstrapped training data sets sampled from each training data set. Gaussian kernel function is utilized in examples, which is

$$K(x_1, x_2) = \exp(-\frac{1}{\sigma^2}||x_1 - x_2||^2).$$

To illustrate the estimation performance of the SVQR ensemble, we compare it with the single SVQR via 100 training data sets, where the root mean squared error (RMSE) is used as the estimation performance measure defined by

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^{n} (\widehat{q}_\theta(x_i) - q_\theta(x_i))^2 \right)^{1/2} \quad \text{for} \quad \theta = 0.1, 0.5, 0.9.$$

Example 4.1 We generate training data which include the homoscedastic structure of error terms: $x_i \sim U(0,1)$, $y_i = \mu(x_i) + 0.15\epsilon_i$ with $\mu(x_i) = \sin(\pi x_i)$ and $\epsilon_i \sim N(0,1)$ for $i = 1, \cdots, 100$. Figure 4.1 shows a family of true quantile regression functions (solid lines) and quantile regression functions estimated by the single SVQR (left, dotted lines) and the SVQR ensemble (right, dotted lines) for one training data set. The quantile regression functions for $\theta = 0.1, 0.5, 0.9$ are superimposed on the scatter plots of x and y. As seen from Figure 4.1, in both procedures three estimators of quantile regression functions reflect well the homoscedastic structure of the error term.

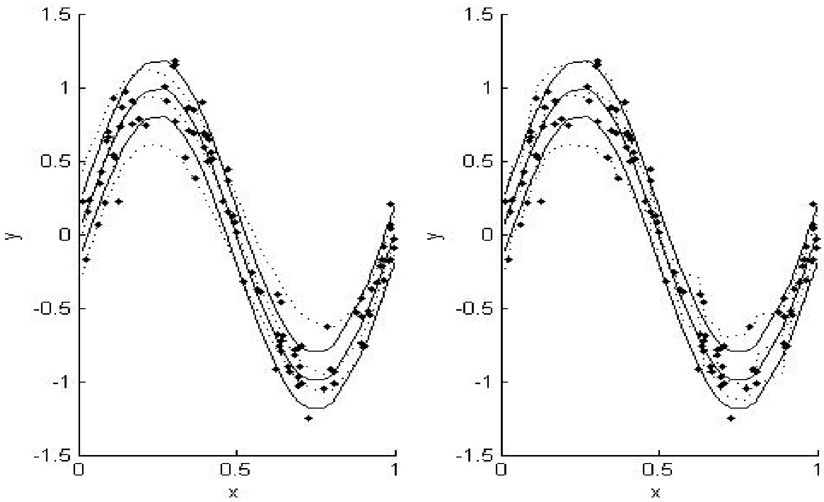

Figure 4.1 An illustration of the true and estimated quantile regression functions by the single SVQR (left) and the SVQR ensemble (right) for a training data set in Example 4.1

Example 4.2 We generate training data which include the heteroscedastic structure of error terms in a similar manner to Shim *et al.* (2009): $x_i \sim U(0,2)$ and $y_i = \mu(x_i) + \sigma(x_i)\epsilon$ with

$\mu(x_i) = \sin(2\pi x_i)$, $\sigma(x_i) = \sqrt{\frac{2.1-x_i}{4}}$ and $\epsilon \sim \chi_2^{(2)} - 2$ for $i = 1, \cdots, 100$. Here $\chi_2^{(2)}$ is the chi-squared distribution with degree of freedom 2. Figure 4.2 shows a family of true quantile regression functions (solid lines) and quantile regression functions estimated by the single SVQR (left, dotted lines) and the SVQR ensemble (right, dotted lines) for one training data set. The quantile regression functions for $\theta = 0.1, 0.5, 0.9$ are superimposed on the scatter plots of x and y. As seen from Figure 4.2, in both procedures three estimators of quantile regression functions reflect well the heteroscedastic structure of the error term but for θ =0.9 the SVQR ensemble looks to reflect better at x values near 0. They have their (local) minima and (local) maxima at different x values. For example, the 0.1th, 0.5th and 0.9th quantile regression functions have minima at x =0.74, 0.74 and 1.78, respectively, and maxima at 1.25, 1.25 and 0.24, respectively.

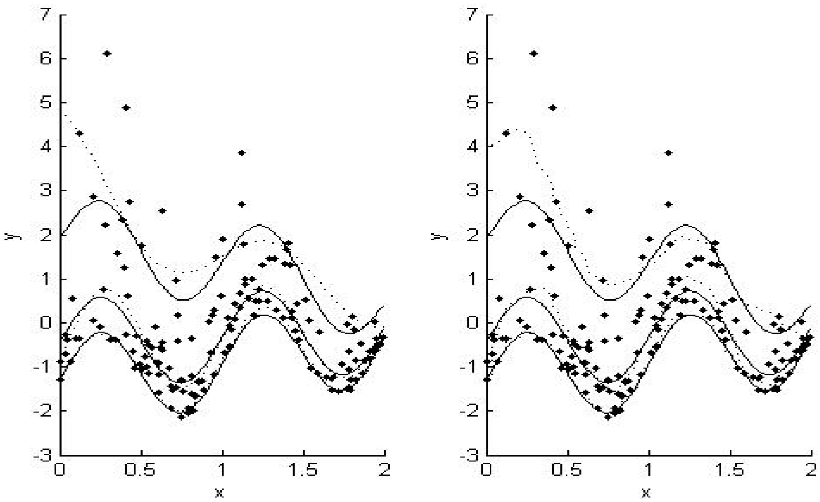

Figure 4.2 An illustration of the true and estimated quantile regression functions by the single SVQR (left) and the SVQR ensemble (right) for a training data set in Example 4.2

Table 4.1 shows the averages of 100 RMSEs from the single SVQR and the SVQR ensemble in Example 4.1 (left) and 4.2 (right). From table we can see that the SVQR ensemble shows smaller RMSE's than the single SVQR, which implies that the SVQR ensemble provides better estimation performance.

Table 4.1 Average of RMSEs of the quantile model using the single SVQR and the SVQR ensemble in Example 4.1 (left) and Example 4.2 (right) (standard error in parenthesis)

θ	SVQR	SVQR ensemble	SVQR	SVQR ensemble
0.1	0.1298 (0.0033)	0.0837 (0.0019)	0.3903 (0.0039)	0.3190 (0.0047)
0.5	0.0692 (0.0013	0.0621 (0.0012)	0.3545 (0.0072)	0.2795 (0.0089)
0.9	0.1307 (0.0031)	0.0827 (0.0018)	0.6962 (0.0186)	0.6748 (0.0215)

5. Conclusions

In this paper, we dealt with estimating the quantile regression function by SVQR ensemble with bagging, where the weights used in aggregating are estimated by minimizing the proposed penalized objective function. Through the example we found that the proposed method derives the good estimation ability. We also found that the model selection of SVQR ensemble with bagging should be improved by incorporating an alternative to the quadratic programming problem in SVQR, which will be studied in future work.

References

Breiman, L. (1996). Bagging predictors. *Machine Learning*, **24**, 123-140.

Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap*, Chapman & Hall/CRC, Boca Raton.

Hwang, C. (2010). M-quantile regression using kernel machine technique. *Journal of the Korean Data & Information Science Society*, **21**, 973-981.

Hwang, C. and Shim, J. (2005). A simple quantile regression via support vector machine. *Lecture Notes in Computer Science*, **3610**, 512-520.

Hwang, C. and Shim, J. (2012). Smoothing Kaplan-Meier estimate using monotone suppot vector regression. *Journal of the Korean Data & Information Science Society*, **23**, 1045-1054.

Kim, H., Pang, S., Je, H., Kim D. and Bang, S. Y. (2003). Constructing support vector machine ensemble. *Pattern Recognition*, **36**, 2357-2767.

Koenker, R. (2005). *Quantile regression*, Cambridge University Press, Cambridge.

Koenker, R. and Bassett, G. (1978). Regression quantile. *Econometrica*, **46**, 33- 50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, **40**, 122-142.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*, University of California Press, Berkeley, 481- 492.

Lee, J. and Shim, J. (2010). Restricted support vector regression without crossing. *Journal of the Korean Data & Information Science Society*, **21**, 319-325.

Li, Y., Liu, Y. and Zhu, J. (2007). Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, **102**, 255-268.

Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A*, 415-446.

Saunders, C., Gammerman, A. and Vovk, V. (1998). Ridge regression learning algorithm in dual variables. In *Proceedings of the 15th International Conference on Machine Learning*, 515-521.

Schapire, R., Freund, Y., Bartlett, P. and Lee, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *The Annals of Statistics*, **26**, 824-832.

Shim, J. and Hwang, C. (2012). M-quantile kernel regression for small area estimation. *Journal of the Korean Data & Information Science Society*, **23**, 749-756.

Shim, J. and Hwang, C. (2013). Expected shortfall estimation using kernel machines *Journal of the Korean Data & Information Science Society*, **24**, 625-636.

Shim, J., Seok, K. H. and Hwang, C. (2009). Non-crossing quantile regression via doubly penalized kernel machine. *Computational Statistics*, **24**, 83-94.

Takeuchi, I., Le, Q. V., Sears, T. D. and Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, **7**, 1231-1264.

Vapnik, V. N. (1995). *The nature of statistical learning theory,* Springer, New York.

Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: Applications and current research area. *The Statistician*, **52**, 331-350.

Yuan, M. (2006). GACV for quantile smoothing splines. *Computational Statistics & Data Analysis*, **50**, 813-829.