Support vector quantile regression using asymmetric e-insensitive loss function

Kyung Ha Seok, Daehyeon Cho
Department of Data Science and Institute of Statistical Information Inje University
Kimhae, 621-749, Korea
Email:statskh@inje.ac.kr, statcho@inje.ac.kr

Changha Hwang
Department of Statistics
Dankook University
Yongin-si, Gyeonggi-do, 448-701, Korea
Email: chwang@dankook.ac.kr

Jooyong Shim
Department of Applied Statistics
Catholic University of Daegu
Gyungbuk, 702-701, Korea
Email: ds1631@hanmail.net

***Abstract*- Support vector quantile regression (SVQR) is capable of providing a good description of the linear and nonlinear relationships among random variables. In this paper we propose a sparse SVQR to overcome a weak point of SVQR, nonsparsity. The asymmetric e-insensitive loss function is used to efficiently provide the sparsity Experimental results are then presented; these results illustrate the performance of the proposed method by comparing it with nonsparse SVQR.**

Keywords : quantile regression; sparsity ; support vectors; support vector quantile regression

I. INTRODUCTION

Quantile regression has been a popular method for estimating the quantiles of a conditional distribution on the values of covariates since [1] introduced linear quantile regression. 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 to, and look at current research areas of quantile regression can be found in [12] and [2]. Support vector machine (SVM) is used as a new technique for regression and classification problems. The SVM is based on the structural risk minimization (SRM) principle, which has been shown to be superior to the traditional empirical risk minimization (ERM) principle. SRM minimizes an upper bound on the expected risk, unlike ERM, which minimizes the error on the training data. By minimizing this bound, high generalization performance can be achieved. In particular, for the SVM regression case, SRM results in regularized ERM with e-insensitive loss function. Introductions to and overviews of recent developments of SVM can be found in [9], [10], [6] and [11]

Sparsity is known as an important feature of kernel regression models. It provides efficiency in predicting the regression function. SVM provides sparsity in which the number of support vectors depends on the number of training data. A small number of support vectors implies sparsity of the model. [8] proposed a Bayesian approach referred to as the relevance vector machine (RVM), providing more sparsity. However. RVM has computational problems since there are no closed-form solutions for maximizing the marginal likelihood. Support vector quantile regression (SVQR) can be obtained by applying SVR with a check function instead of an e-insensitive loss function into the quantile regression [7]. But SVQR does not provide sparsity due to zero insensitiveness of check function. Here we define support vectors as the index numbers corresponding to nonzero Lagrange multiplier differences. By using an asymmetric e-insensitive loss function we can take the support vectors efficiently depending on the value of quantiles. In this paper we use an asymmetric e-insensitive loss function in SVQR to provide the sparsity. The proposed loss function is designed to provide more sparsity by adjusting insensitiveness according the sign of residuals. In Section 2 we propose a sparse SVQR using an asymmetric e-insensitive loss function and perform numerical studies through examples. In Section 3 we give the conclusions.

II. SUPPORT VECTOR QUANTILE REGRESSION

A. SVQR

Let the training data set D be denoted by (x_i, y_i), $i = 1, \dots, n$ with each input vector $x \in R^d$ and the output $y_i \in R$, which is linearly or nonlinearly related to the input vector 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, $\phi(x_i)'\phi(x_j) = K(x_i, x_j)$ ([4]). Several choices of the kernel $K(\cdot,\cdot)$ are possible. We consider the nonlinear case, in which the θth quantile function, given x, $q_\theta(x)$ for $\theta \in (0,1)$, can be regarded as a nonlinear function of input vector x.

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

$$\min \ \frac{1}{2}w'w + C\sum_{i=1}^{n} h\big(y_i - q_\theta(x_i)\big) \qquad (1)$$

where $h(r) = \theta r I(r > 0) + (\theta - 1) r I(x \le 0)$ for $\theta \in (0,1)$, where $I(\cdot)$ is the indicated function. We can express the quantile regression problem by formulation for SVM as follows.

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

subject to

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

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

$$L = \frac{1}{2}w'w + C\sum_{i=1}^{n} (\theta\xi_i + (1-\theta)\xi_i^*) \qquad (3)$$
$$-\sum_{i=1}^{n} \alpha_i(\xi_i - (y_i - w'\phi(x_i) - b))$$
$$-\sum_{i=1}^{n} \alpha_i^*(\xi_i - (w'\phi(x_i) - b - y_i))$$
$$-\sum_{i=1}^{n} \eta_i\xi_i - \sum_{i=1}^{n} \eta_i * \xi_i *$$

We notice that the non-negative constraints $\alpha_i, \eta_i, \alpha_i^*, \eta_i^* \ge 0$ should be satisfied. After taking partial derivatives of (3) with regard to the primal variables (w, ξ_i, b) and plugging them into (3), we have the optimization problem below.

$$\max \ L = -\frac{1}{2}\sum_{i,j=1}^{n} (\alpha_i - \alpha_j)(\alpha_i^* - \alpha_j^*)K(x_i, x_j) \qquad (4)$$
$$+\sum_{i=1}^{n} (\alpha_i - \alpha_i^*)y_i$$

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

Solving the above problem with the constraints determines the optimal Lagrange multipliers $\hat{\alpha}_i$ and $\hat{\alpha}^*$. Thus, the estimated θth quantile function given the input vector x_0 is obtained as

$$\hat{q}_\theta(x_0) = K(x_0, x)(\hat{\alpha} - \hat{\alpha}^*) + \hat{b} \qquad (5)$$

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

$$\hat{b} = \frac{1}{n_s}\sum_{i=1}^{n_s} (y_i - (\hat{\alpha} - \hat{\alpha}^*)'K_i)$$

where $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_n)'$, $\hat{\alpha}^* = (\hat{\alpha}_1^*, \ldots, \hat{\alpha}_n^*)'$ and n_s is the size of the set $\{i = 1, \cdots, n | C(\theta - 1) < \alpha_i - \alpha_i^* < C\theta\}$ and K_i is the ith row of the kernel matrix $K = \{K(x_i, x_j)\}_{n\times n}$

In the nonlinear case, w is no longer explicitly given. However, it is uniquely defined in the weak sense by the dot products. Here the linear regression model can be regarded as a special case of the nonlinear regression model by using identity feature mapping function, that is, $\phi(x) = x$ which implies the linear kernel such that $K(x_1, x_2) = x'_1 x_2$.

B. Sparse SVQR

With an asymmetric e-insensitive loss function $h(\cdot)$, shown in Fig. 1, the θth quantile function can be defined as a function of any solution to the optimization problem,

$$\min \frac{1}{2}w'w + C\sum_{i=1}^{n} h\big(y_i - q_\theta(x_i)\big) \qquad (6)$$

where $h(r) = 0$ if $\frac{\theta}{\theta-1}e \le r \le \frac{1-\theta}{\theta e}$, $h(r) = \theta r - (1-\theta)e$ if $r > \frac{1-\theta}{\theta}e$ and $h(r) = (\theta - 1)r - \theta e$ if $r < \frac{\theta}{\theta-1}e$ for $\theta \in (0,1)$. The check function used in nonsparse SVQR and the asymmetric e-insensitive loss function are illustrated in Fig. 1 with $\theta = 0.35,\ 0.75$ and $e = 0.2$.

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

$$\min \ \frac{1}{2}w^{'}w + C\sum_{i=1}^{n} (\theta\xi_i + (1-\theta)\xi_i^*) \qquad (7)$$

subject to

$$y_i - w'\phi(x_i) - b \le \xi_i + 1 - \theta\theta e,$$
$$w'\phi(x_i) + b - y_i \le \xi_i^* + \theta 1 - \theta e$$
$$\xi_i \ge 0, \xi_i^* \ge 0,$$

where $e > 0$ and C is a regularization parameter penalizing the training errors. We construct a Lagrange function as follows:

$$L = \frac{1}{2}w^{'}w + C\sum_{i=1}^{n} (\theta\xi_i + (1-\theta)\xi_i^*) \qquad (8)$$
$$-\sum_{i=1}^{n} \alpha_i(\xi_i + \frac{1-\theta}{\theta}e - (y_i - w'\phi(x_i) - b))$$
$$-\sum_{i=1}^{n} \alpha_i^*(\xi_i^* + \frac{\theta}{1-\theta}e - (w'\phi(x_i) + b - y_i))$$
$$-\sum_{i=1}^{n} \eta_i\xi_i - \sum_{i=1}^{n} \eta_i^*\xi_i^*$$

We notice that the non-negative constraints $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \ge 0$ should be satisfied. After taking partial derivatives of (8) with regard to the primal variables (w, ξ_i, b) and plugging them into (8), we have the optimization problem below.

$$\min L = \frac{1}{2}\sum_{i,j=1}^{n} (\alpha_i - \alpha_j)(\alpha_i^* - \alpha_j^*)K(x_i, x_j) \qquad (9)$$
$$+\sum_{i=1}^{n} \alpha_i(\frac{1-\theta}{\theta}e - y_i) + \sum_{i=1}^{n} \alpha_i^*(\frac{\theta}{1-\theta}e + y_i)$$

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

Solving the above problem with the constraints determines the optimal Lagrange multipliers α_i and α_i^*. Here the input vector x_i, corresponding to positive α_i or α_i^*, is called the support vector. Thus, the estimated θth quantile function, given the input vector x_0, is obtained as (5)

$$\hat{q}_\theta(x_0) = K(x_0, x)(\hat{\alpha} - \hat{\alpha}^*) + \hat{b}$$

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

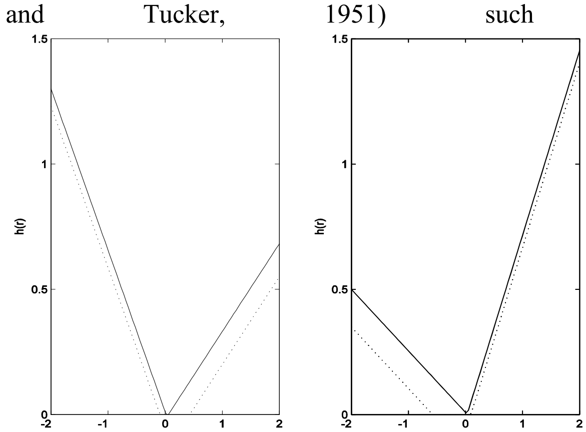

Figure 1: The check function(solid line) and the asymmetric 0.2-insensitive loss function(dotted line) with $\theta = 0.35$(Left) and $\theta = 0.75$(Right).

$$\hat{b} = \frac{1}{(n_1+n_2)}\left\{\sum_{i\in I_1}\left(y_i - (\hat{\alpha}-\hat{\alpha}^*)' K_i - \frac{1-\theta}{\theta}e\right) + \sum_{i\in I_2}\left(y_i - (\hat{\alpha}-\hat{\alpha}^*)' K_i + \frac{\theta}{1-\theta}e\right)\right\}$$

where n_1 is the size of the set $I_1 = \{i = 1,\cdots,n | 0 < \hat{\alpha}_i < C\theta\}$ and n_2 is the size of the set $I_2 = \{i = 1,\cdots,n | 0 < \hat{\alpha}_i^* < C(1-\theta)\}$.

We can see that $\{i = 1,\cdots,n | 0 < \hat{\alpha}_i^* \le (1-\theta)C\} = \{i = 1,\cdots,n | y_i \ge \hat{q}_\theta(x_i) + \frac{1-\theta}{\theta}e\}$ and $\{i = 1,\cdots,n | 0 < \hat{\alpha}_i \le \theta C = \{i = 1,\cdots,n | y_i \le \hat{q}_\theta(x_i) - \frac{\theta}{1-\theta}e\}$, which are indices of data points with support vectors that are not in the asymmetric e-tube.

C. Numerical Studies

We illustrate the performance of the sparse quantile regression estimation through the simulated data for nonlinear regression cases.

100 data sets are generated to present the prediction performance of the proposed method. Each data set consists of 100 x's and 100 y's. Here x's are equally spaced ranging from 0 to π; y's are generated from a normal distribution $N(1+sin(x),0.1)$. The true θth quantile function is given as

$$q_\theta(x) = 1 + sin(x) + 0.1\Phi^{-1}(\theta) \;\; for\; \theta \in (0,1),$$

where $\Phi(\cdot)$ is the cdf of $N(0,1)$ distribution. The radial basis kernel function is utilized in this example, which is

$$K(x_1, x_2) = exp(-(x_1 - x_2)^2/\sigma^2).$$

The hyperparameters (e, C, σ^2) were chosen as (0.05, 100,1) by 5-fold cross-validation. Fig. 2 shows the estimated θth quantile regression functions imposed on the scatter plots of one data set for $\theta = 0.1$(Left) and $\theta = 0.9$(Right). From Fig. 2 we can see that the proposed method provides the sparsity.

From 100 data sets we obtain a mean squared error of $q_\theta(x) - \hat{q}_\theta(x)$ to compare the performance of sparse SVQR to SVQR, with results shown in Table 1. From the table we can see that both give similar estimation performance but that the proposed sparse SVQR provides sparsity.

TABLE 1: THE AVERAGE OF 100 MSEs OF $q_\theta(x) - \hat{q}_\theta(x)$ FOR $\theta = 0.1$, 0.5 AND 0.9 (STANDARD DEVIATIONS OF MSEs IN PARENTHESIS)

θ	Sparse SVQR	nonsparse SVQR
0.1	0.0206(0.0102)	0.0207(0.0106)
0.5	0.0105(0.0056)	0.0120(0.0062)
0.9	0.0202(0.0108)	0.0176(0.0081)

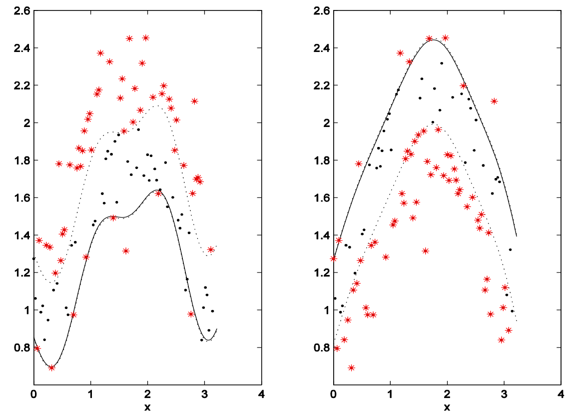

Figure 2: Estimated θth quantile regression functions $(\hat{q}_\theta(x))$ imposed on the scatter plots of 100 data points of a data set. dots=data points with nonsupport vector, stars=data points with support vectors, solid line = $\hat{q}_\theta(x)$, dotted line(upper) $=\hat{q}_\theta(x) + \frac{1-\theta}{\theta}e$ and dotted line(lower) = $\hat{q}_\theta(x) + \frac{\theta}{1-\theta}e$.

III. CONCLUSION

In this paper, we dealt with estimating the quantile regression function by SVQR using an asymmetric e-insensitive loss function. Through the example we showed that the proposed method provides sparsity and performance similar to that of SVQR. The model selection method, such as the generalized approximate cross-validation function, will be studied in our further research using the effective dimensionality of the fitted model.

ACKNOWLEDGMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (2009-0072369).

REFERENCES

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

[2] Koenker, R. (2005). Quantile Regression. Cambridge University Press.

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

[4] 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.

[5] Shim, J. and Hwang, C. (2009). Support vector censored quantile regression under random censoring. Computational Statistics and Data Analysis 53, 4, 912-917.

[6] Smola, A. and Scholkopf, B. (1998). On a Kernel-based method for pattern recognition, regression, approximation and operator inversion. Algorithmica, 22, 211-231.

[7] Takeuchi, I., Le, Q. V., Sears, T. D. and Smola, A. J.(2006).

Nonparametric quantile estimation. Journal of Machine Learning Research, 7, 1231-1264.

[8] Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research, 1, 211-244.

[9] Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. Springer, New York.

[10] Vapnik, V. N. (1998). Statistical Learning Theory. John Wiley, New York.

[11] Wang, L.(Ed.) (2005). Support Vector Machines: Theory and Application. Springer, Berlin Heidelberg New York.

[12] Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: applications and current research area. The Statistician, 52(3), 331-350.