A Weighted Twin Support Vector Regression

Smit Lunagariya , Guided by Dr. Debdas Ghosh

December 3, 2019

1.Abstract

Twin Support Vector Regression is a new regression algorithm, which aims at finding ε-sensitive up and down-bound functions for the training points. In order to do so, one needs to resolve a pair of smaller-sized quadratic programming problems (QPPs) rather than a single large one in a classical SVR. However, the same penalties are given to the samples of TSVR. Then, upon this algorithm, a new algorithm was proposed called as weighted twin support vector regression, wherein the samples in the different positions are proposed to give different penalties. The final Regressor can avoid the over fitting problem to a certain extent and yield great generalization ability.

2.Introduction

Support Vector Machine made the state of art and one of the most used classifiers. SVM has many advantages such as when it solves a QPP, it assures that a optimal solution is obtained, it is unique solution. Secondly, it derives its sparse and robust solution by maximising the margin between the two classes.

However, one of the main challenges of the standard SVM is the high computational complexity which is of the order of n^3, where n is the total size of the training data. In order to improve its computational speed, Twin Support Vector machine was proposed.

A Twin Support Vector machine generates two non parallel hyperplanes by solving two smaller sized QPPs such that each hyperplane is closer to one and as far as possible from the other class. This approach made the learing speed much faster approximated as four times that of SVM. In TSVR, same penalties are given to the samples. However as samples located in different positions, it is more resonable to give different penalties to them.

Inspired from the studies of TSVR, weighted TSVR was proposed in which two weighted coefficients σ_1 and σ_2 into the TSVR were introduced. By dividing the whole plane into different parts, and bringing different penalties to the samples depending on their different positions.

3.Support vector Regression

Consider the training set T=$\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$ and let the matrix A=$(x_1,x_2,...,x_n)$ and the matrix Y=$(y_1,y_2,...,y_n)$.

The non-linear SVR seeks to find a regression function in a high dimensional feature space tolerating the small error in fitting the given dataset. This can be achieved by utilizing the ε-insensitive loss functions that sets an ε-insensitive tube around the data, within which errors are discard. The nonlinear SVR can be obtained by resolving the following QPP:

$$\min_{w,b,\xi,\xi^*} \ \frac{1}{2}||w||^2 + c.(e^T\xi + e^T\xi^*) \qquad (1)$$

$$s.t. \ (\varphi^T(A)w + eb) - Y \leq e\varepsilon + \xi,$$

$$Y - (\varphi^T(A)w + eb) \leq e\varepsilon + \xi^*,$$

$$\xi \geq 0e, \ \ \xi^* \geq 0e,$$

where c is a parameter chosen priori, which tradeoff between the fitting errors and flatness of the regression functions, ξ and $\xi^*$are the slack vectors, and e is the vector of ones of appropiate dimensions.

By introducing the Lagrangian multipiers α and α^*, we can derive the dual problem of the QPP as follows:

$$\max_{\alpha,\alpha^*} \ -\frac{1}{2}(\alpha^*-\alpha)^T K(A,A)^T(\alpha-\alpha^*) + Y^T(\alpha-\alpha^*) + \varepsilon e^T(\alpha+\alpha^*), \qquad (2)$$

$$s.t. \ e^T(\alpha-\alpha^*) = 0,$$

$$0e \leq \alpha, \ \alpha^* \leq ce.$$

Upon solving the QPP we obtain the regression function as,

$$f(x) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)K(x_i,x) + b \qquad (3)$$

Here $K(x_i,x) = (\varphi(x_i).\varphi(x))$ represents the kernel function which is the dot product of high dimensional feature space.

4.Twin Support Vector Regression

In order to improve the computational efficiency of the SVM, an efficient TSVR was proposed that generates two ε-insensitive functions and the final regressor is decided by the mean of both the functions:

$$f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2}(w_1 + w_2) + \frac{1}{2}(b_1 + b_2) \tag{4}$$

The TSVR resolves the following pair of QPPs:

$$\underset{w_1,b_1,\xi}{min} \quad \frac{1}{2}||Y - e\varepsilon_1 - K(A, A^T)w_1 + eb_1||^2 + c_1 e^T \xi, \tag{5}$$

$$s.t. \;\; Y - (K(A, A^T)w_1 + eb_1) \geq e\varepsilon_1 - \xi, \quad \xi \geq 0e,$$

and

$$\underset{w_2,b_2,\eta}{min} \quad \frac{1}{2}||Y + e\varepsilon_2 - K(A, A^T)w_2 + eb_2||^2 + c_1 e^T \eta, \tag{6}$$

$$s.t. \;\; (K(A, A^T)w_2 + eb_2) - Y \geq e\varepsilon_2 - \eta, \quad \eta \geq 0e$$

Similarly, the dual problems of the QPPs are as the follows:

$$\underset{\alpha}{max} \quad -\frac{1}{2}\alpha^T H(H^T H)^{-1} H^T \alpha + f^T H(H^T H)^{-1} H^T \alpha - f^T \alpha \tag{7}$$

$$s.t. \quad 0e \leq \alpha \leq c_1 e$$

and

$$\underset{\beta}{max} \quad -\frac{1}{2}\beta^T H(H^T H)^{-1} H^T \beta - h^T H(H^T H)^{-1} H^T \beta + h^T \beta, \tag{8}$$

$$s.t. \quad 0e \leq \beta \leq c_2 e$$

where $for\ nonlinear\ case\ H = [K(A, A^T)\ e] \quad and\ for\ linear\ case\ H = [A\ e], \quad f = Y - e\epsilon_1, \quad h = Y + e\epsilon_2.$

Once the dual QPPs are solved we can get,

$$[w_1\ b_1]^T = (H^T H)^{-1} H^T (f - \alpha), \tag{9}$$

$$[w_2\ b_2]^T = (H^T H)^{-1} H^T (h + \beta) \tag{10}$$

The TSVR is comprised of pair of QPPs such that each of them determines one of the up- or down-bound functions.

5.Weighted Twin Support Vector Regression

In TSVR, the same penalties are given to the samples. This implies that no matter where the samples stay, they will suffer from the same penalties. in fact, they have different effects on the bound function.

In the feature space, a linear regression function is implemented which corresponds to a nonlinear regression function in the input space.

$$f_1(x) = K(A, x^T)w_1 + b_{1,} \quad f_2(x) = K(A, x^T)w_2 + b_2 \tag{11}$$

where K is any arbitary kernel function. Similarly, the weighted TSVR formulation of the non-linear case is of the form

$$\min_{w_1,b_1,\xi,\xi^*} \frac{1}{2}||Y - e\epsilon_1 - K(A, A^T)w_1 + eb_1||^2 + c_1(e^T\xi + \sigma_1 e^T\xi^*), \tag{12}$$

$$s.t.\ \ Y - (K(A, A^T)w_1 + eb_1) \geq e\epsilon_1 - \xi - \xi^*,$$

$$0e \leq \xi \leq \epsilon_1 e, \quad \xi^* \geq 0e,$$

and

$$\min_{w_2,b_2,\eta,\eta^*} \frac{1}{2}||Y + e\epsilon_2 - K(A, A^T)w_2 + eb_2||^2 + c_2(e^T\eta + \sigma_2 e^T\eta^*), \tag{13}$$

$$s.t.\ \ (K(A, A^T)w_2 + eb_2) - Y \geq e\epsilon_2 - \eta - \eta^*,$$

$$0e \leq \eta \leq \epsilon_2 e, \quad \eta^* \geq 0e.$$

where parameters $c_1, c_2 > 0$ and $\epsilon_1, \epsilon_2 > 0$. ξ, ξ^*, η, η^* are slack vectors. Penalty parameters $\sigma_1, \sigma_2 > 1$,which implies that the larger penalties are given to the samples locating far off from the margin.

To resolve the QPP, we use the Langrangian function with $\alpha, \lambda, \gamma, \phi$ as the Lagrangian multipliers, from which derive the dual problem of the QPP as follows:

$$\max_{\alpha,\gamma} \ \ -\frac{1}{2}\alpha^T H(H^T H)^{-1}H^T\alpha + f^T H(H^T H)^{-1}H^T\alpha - f^T\alpha - \epsilon_1 e^T\gamma \tag{14}$$

$$s.t. \quad 0e \leq \alpha \leq c_1\sigma_1, \quad \gamma \geq 0e,$$

Similarly,

$$\max_{\beta,\lambda} \quad -\frac{1}{2}\beta^T H(H^T H)^{-1}H^T\beta - h^T H(H^T H)^{-1}H^T\beta + h^T\beta - \epsilon_2 e^T\lambda \tag{15}$$

$$s.t. \quad 0e \leq \beta \leq c_2\sigma_2, \quad \lambda \geq 0e$$

where *for nonlinear case* $H = [K(A, A^T)\ e]$ *and for linear case* $H = [A\ e], \quad f = Y - e\epsilon_1, \quad h = Y + e\epsilon_2$.

Once the dual QPPs are solved we can get the vectors as,

$$[w_1\ b_1]^T = (H^T H + \delta I)^{-1}H^T(f - \alpha), \tag{16}$$

$$[w_2\ b_2]^T = (H^T H + \delta I)^{-1}H^T(h + \beta) \tag{17}$$

where δ is the regularization parameter.

Finally, we can achieve the end regressor as follows:

$$f(x) = \frac{1}{2}(f_1(x) + f_2(x)) = \frac{1}{2}(w_1 + w_2)^T K(A, x^T) + \frac{1}{2}(b_1 + b_2) \tag{18}$$

6.Programming in Python

6.1 Python-cvxopt library

The function 'qp' is an interface to convex programming of quadratic programs. It also provides the option of using the quadratic programming solver from MOSEK

syntax:

$cvxopt.solvers.qp(P, q[, G, h[, A, b[, solver[, initvals]]]])$

This solves the pair of primal and dual convex quadratic programs

$$min \ \frac{1}{2}x^T Px + q^T x \tag{19}$$

$$subject\,to \quad Gx \leq h,\ Ax = b$$

The inequalities are componentwise vector inequalities.

Example:

Consider the dual problem from twin support vector regression

$$\max_{\alpha} \quad -\frac{1}{2}\alpha^T H(H^T H)^{-1}H^T\alpha + f^T H(H^T H)^{-1}H^T\alpha - f^T\alpha \tag{20}$$

$$s.t. \quad 0e \leq \alpha \leq c_1 e$$

This is equivalent to,

$$\underset{\alpha}{min} \quad \frac{1}{2}\alpha^T H(H^T H)^{-1}H^T\alpha - f^T H(H^T H)^{-1}H^T\alpha + f^T\alpha \tag{21}$$

$$s.t. \quad 0e \leq \alpha \leq c_1 e$$

Hence from (19) and (21)

$$P = H(H^T H)^{-1}H^T \tag{22}$$

$$q = f^T - f^T H(H^T H)^{-1}H^T \tag{23}$$

$$G = verticalstack(identity\, matrix * (-1), identity\, matrix) \tag{24}$$

$$h = verticalstack(column\, matrix(0), column\, matrix(c1)) \tag{25}$$

$$A = 0, b = 0 \tag{26}$$

Therefore using the above relation, the cvxopt library is used to solve the dual problem obtained the algorithms.

7.Numerical experiments

7.1 Evaluation Criteria

MAE: Mean absolute error , defined as ,

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y_i}| \tag{27}$$

RMSE: Root mean squared error, defined as,

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y_i})^2} \tag{28}$$

SSE: Sum of squared error of testing, defined as,

$$SSE = \sum_{i=1}^{m}(y_i - \hat{y_i})^2 \tag{29}$$

SST: Sum of squared deviation of testing samples, defined as,

$$SST = \sum_{i=1}^{m} (y_i - \bar{y_i})^2 \tag{30}$$

SSR: Sum of squared deviation that can be explained by the estimator is defined as,

$$SSR = \sum_{i=1}^{m} (\hat{y}_i - \bar{y_i})^2 \tag{31}$$

7.2 Experiment on artificial dataset

In this arttifical experiment on the regression of the Sinc function. Sinc function is defined as,

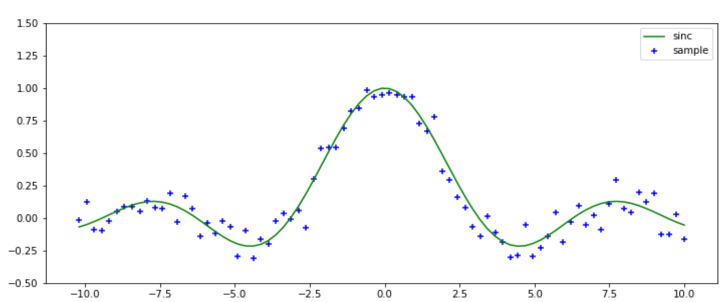
$$y = Sinc(x) = \frac{\sin(x)}{x} \tag{32}$$

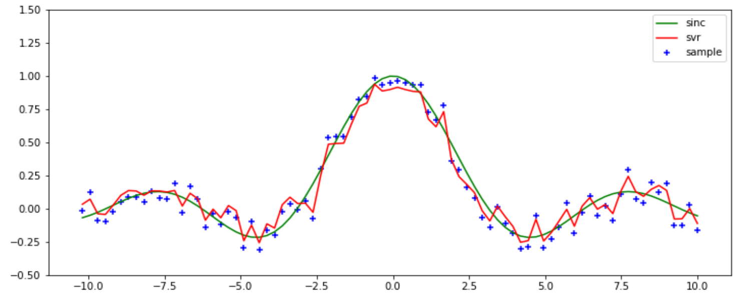
The gererated 161 points are preturbed by Gaussian noises with zero mean and variance of 0.1^2. Specifically, we have the following training samples:

$$y_i = \frac{\sin(x_i)}{x_i} + \xi_i, \quad \xi_i \sim N(0, 0.1^2) \tag{33}$$

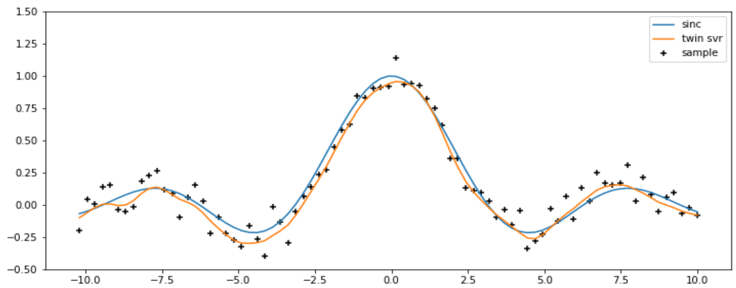
The generated 161 points are split into two subsets, one of which includes 81 pointd, as training set. The other subset includes 80 points, as testing set. The performance comparisons of the three algorithms are summarized in the below Table.

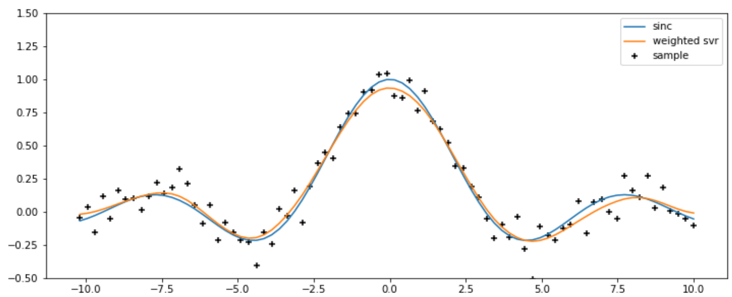
Algorithm	Parameters(c, ϵ, p, σ)	MAE	RMSE	SSE	SSR	SST	SSR/SST	SSE/SST
SVR	(512,0.05,0.25,0)	0.068	0.082	0.542	8.490	9.943	0.854	0.055
Twin SVR	(0.1,10,0.11,0)	0.047	0.055	0.248	9.785	9.942	0.984	0.025
Weighted SVR	(0.1,10,0.11,0.1)	0.029	0.034	0.092	8.974	9.942	0.903	0.009





1





8.Future Work

The algorithms can be futher extended for the fuzzy numbers. A fuzzy number is a generalization of a regular real number in the sense that t does not refer to one single value but rather to a connected set of possible values, where each possible value has its own weight between 0 and 1.

9.References

[1] Y. Xu,L. Wang-Paper on Weighted Support Vector Regression.College of Science, China Agricultural University, China.

[2]-S K Shevade- Computer Numerical Optimization(nptel).IISc Bangalore, India.