ANALYSIS OF DATA PERTURBATION DUE TO DATA ADDITION OR MISSING IN LINEAR SUPPORT VECTOR REGRESSION

Pei Wang^a, Chun Cai^b, Dalian Liu^a and Xinfeng Wang^a

^a Department of Basic Courses, Beijing Union University, Beijing, 100101, P. R. China

^b College of Applied Arts and Science, Beijing Union University, Beijing, 100191, P. R. China

Abstract

The support vector regression based on the Statistical Learning Theory avoids the inadequacy of traditional function approximation methods and proves to be highly available for learning and generalization. Its mathematical model is a convex quadratic programming, of which the optimality condition (i.e., the KKT condition) is sufficient and necessary for an optimal solution. In this paper, the quadratic programming of the support vector regression is reduced to linear programming. Based on sensitivity analysis of the linear programming, analysis is conducted on data perturbation due to the decrease (missing) and increase of the initial data. This paper provides the sufficient conditions for remaining the support vectors unchanged under the above cases of perturbation and explains the change in the fitting function.

*Corresponding author.

E-mail address: ldtxinfeng@buu.edu.cn (Xinfeng Wang).

Copyright © 2018 Scientific Advances Publishers
2010 Mathematics Subject Classification: 68W40, 68Q25.
Submitted by Jianqiang Gao.
Received March 30, 2018

Keywords: support vector regression, linear programming, sensitivity analysis, data perturbation, KKT condition.

1. Introduction

The support vector machine is a learning approach based on the Statistical Learning Theory [1], which proves to be highly available for learning and generalization. It mainly deals with three categories of problems: mode identification (classification), regression estimate (curve fitting) and density estimate. Particularly, the support vector machine dealing with classification is called the support vector classification while that dealing with regression is called the support vector regression (SVR).

The support vector machine is a convex quadratic programming, thus the optimal solution is the global optimal solution, and the optimality condition (i.e., the KKT condition) is both sufficient and necessary for reaching an optimal solution. However, we usually do not solve the said convex quadratic programming directly. Instead, we convert it to an equivalent dual problem via the KKT condition and then solve it. The advantages of using dual problems are stated as follows: firstly, it is still a convex quadratic programming; secondly, in a dual problem, the Lagrange multiplier exists as an inner product, for which the kernel function [2] can be introduced (to project the inner product to a higher dimensional space), so that the non-linear problem is reduced to a linear problem to solve. As to regression, it is the conversion from non-linear fitting to linear fitting, and the linear fitting function is defined by corresponding Lagrange multiplier.

In this article, the first step is to reduce the quadratic programming of the vector machine to linear programming. The linear programming coefficient changes when certain data points are added or deleted. Thus, we can convert the problem of data perturbation into the problem of

sensitivity analysis of linear programming. Sensitivity analysis of linear programming mainly involves two aspects: firstly, how will the optimal solution of said problem change when one or more parameters of the model change? Secondly, what is the range of change of said parameters when the optimal solution or optimal basis of the problem is to be unchanged? These two aspects exactly correspond to the two problems relating to data perturbation of the support vector regression, i.e., (1) how will the fitting function change when perturbation occurs to a data point? (2) what is the boundary for perturbation so that the fitting function does not change? It is clear that the support vectors in data is a “contributive” vectors for the fitting function, which decides the latter. The support vectors remain unchanged as long as the optimal basis of linear programming remains the same. Therefore, the focus of this paper is laid on the condition for remaining the optimal basis (the support vectors) unchanged if the optimal basis for the linear programming of the support vector machine remains unchanged when certain data points are added or deleted.

2. Programming Model of SVR

2.1. Quadratic programming model and dual problem of SVR

Let $X = \{(x_i, y_i)\}_{i=1}^l$, $x_i \in R^n$, $y_i \in R$, be a training set following the probability density function $P(x, y)$, and $F = \{f \mid f : R^n \rightarrow R\}$ be a function set. The intention is to find a function $f \in F$ so that the expected risk function [3] is minimal, that is, to let $R(f) = \int l(y - f(x)) dp(x, y)$ be the minimum. Here, the loss function $l(y - f(x))$ is defined as follows:

$$l(y - f(x)) = l(|y - f(x, \alpha)|_\varepsilon) = \begin{cases} 0, & \text{if } |y - f(x, \alpha)| \leq \varepsilon, \\ |y - f(x, \alpha)| - \varepsilon, & \text{otherwise.} \end{cases}$$

Such function is called the linear ε -insensitive loss function [3].

Let $f(x) = (\omega \cdot x) + b$ be a linear function. Suppose that all data can be approximated with the linear function within a precision of ε when no error occurs, i.e.,

$$\begin{cases} y_i - (\omega \cdot x_i) - b \leq \varepsilon, \\ (\omega \cdot x_i) + b - y_i \leq \varepsilon. \end{cases} \quad (1)$$

According to Structural Risk Minimization [4], f shall realize a minimal $\|\omega\|^2 + C \sum_{i=1}^l l(|y_i - f(x_i, \alpha)|_\varepsilon)$. In which, $C > 0$ is a constant applied to restrict the degree of punishment on the sample exceeding the limit ε . Introduce slack variable ξ_i, ξ_i^* to establish the programming model of the support vector regression:

$$\begin{aligned} \text{Min} \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y_i - (\omega \cdot x_i) - b \leq \varepsilon + \xi_i \\ & (\omega \cdot x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, l. \end{aligned}$$

The KKT condition is applied to obtain the dual programming of the convex quadratic programming. Introduce the Lagrange multiplier $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ and let

$$\begin{aligned} L(\omega, \xi_i, \xi_i^*) = & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i (-y_i + (\omega \cdot x_i) + b + \varepsilon + \xi_i) \\ & - \sum_{i=1}^l \alpha_i^* (y_i - (\omega \cdot x_i) - b + \varepsilon + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*). \quad (2) \end{aligned}$$

Take the partial derivative with respect to ω , b , ξ_i , ξ_i^* in Equation (2):

$$\begin{cases} \frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0, \\ \frac{\partial L}{\partial b} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0, \\ \frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0, \\ \frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0. \end{cases} \quad (3)$$

The following equation is obtained based on the complementary slackness condition contained in the KKT condition:

$$\begin{cases} \alpha_i (-y_i + (\omega \cdot x_i) + b + \varepsilon + \xi_i) = 0, \\ \alpha_i^* (y_i - (\omega \cdot x_i) - b + \varepsilon + \xi_i^*) = 0, \\ \eta_i \xi_i = 0 \quad \text{or} \quad (C - \alpha_i) \xi_i = 0, \\ \eta_i^* \xi_i^* = 0 \quad \text{or} \quad (C - \alpha_i^*) \xi_i^* = 0. \end{cases} \quad (4)$$

Substitute (3) into (2) and combine (4) to obtain the dual programming:

$$\begin{aligned} \text{Min} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i \cdot x_j) + \varepsilon \sum_{i=1}^l (\alpha_i^* - \alpha_i) - \sum_{i=1}^l (\alpha_i^* - \alpha_i) \\ \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ & 0 \leq \alpha_i^*, \alpha_i \leq C, i = 1, 2, \dots, l. \end{aligned}$$

Solve the programming:

$$\omega = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i. \quad (5)$$

We then obtain the fitting function $f(x) = \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i)(x_i \cdot x) + \bar{b}$, in

which $\bar{\alpha}_i^*$ and $\bar{\alpha}_i$ are the optimal solutions. Compute $\bar{b}$ as follows: selection $\bar{\alpha}_j$ or $\bar{\alpha}_k^*$ from the open interval $(0, C)$. For $\bar{\alpha}_j$, we have $\bar{b} = y_j - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i)(x_i \cdot x_j) + \varepsilon$; for $\bar{\alpha}_k^*$, we can get $\bar{b} = y_k - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i)(x_i \cdot x_k) - \varepsilon$.

Since α_i and α_i^* can not work as the basic variables at the same time, at least one of α_i and α_i^* shall be 0 according to the first two equations in (4). Then, the optimal value obtained from α_i and α_i^* must meet one of the following conditions:

$$\begin{cases} \alpha_i = 0, \alpha_i^* = 0, \end{cases} \quad (6a)$$

$$\begin{cases} \alpha_i = 0, 0 < \alpha_i^* < C, \end{cases} \quad (6b)$$

$$\begin{cases} 0 < \alpha_i < C, \alpha_i^* = 0, \end{cases} \quad (6c)$$

$$\begin{cases} \alpha_i = 0, \alpha_i^* = C, \end{cases} \quad (6d)$$

$$\begin{cases} \alpha_i = C, \alpha_i^* = 0. \end{cases} \quad (6e)$$

According to Equation (5), the vector corresponding to the condition of satisfaction (6a) does not contribute to ω , while the vector corresponding to (6b) ~ (6e) contributes to ω . The vector meeting (6b) ~ (6e) must be a support vector, while a non-support vector must meet (6a). Thus, Equation (5) can also be expressed as $\omega = \sum_{SV-Set} (\alpha_i - \alpha_i^*)x_i$, where

SV-Set stands for the set of support vectors, and the corresponding fitting function can be expressed as $f(x) = \sum_{SV-Set} (\bar{\alpha}_i^* - \bar{\alpha}_i)(x_i \cdot x) + \bar{b}$. It is

clear that the support vectors play a decisive role on the fitting function.

2.2. Linear programming model of SVR

Based on the relationship between the dual problem and solution to the original programming, the original quadratic programming can be reduced to a linear programming:

$$\begin{aligned}
 \text{S-LP:} \quad & \underset{\alpha_i^*, \alpha_i, \xi_i, \xi_i^*, b}{\text{Min}} \quad \frac{1}{l} \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \frac{C}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) \\
 \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) (x_i \cdot x_j) - \xi_j + b \leq \varepsilon + y_j \\
 & - \sum_{i=1}^l (\alpha_i - \alpha_i^*) (x_i \cdot x_j) - \xi_i^* - b \leq \varepsilon - y_j \\
 & \alpha_i, \alpha_i^*, \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, l.
 \end{aligned}$$

Let $\bar{\alpha}_i, \bar{\alpha}_i^*, \bar{\xi}_i, \bar{\xi}_i^*, \bar{b}$ be the solution to the programming, and $f(x) = \sum_{SV-Set} (\bar{\alpha}_i^* - \bar{\alpha}_i) (x_i \cdot x) + \bar{b}$ be the fitting function. Due to the equivalency of the problem, the interpretation for the value of α_i, α_i^* is the same as before. Let $b = b_1 - b_2$, where $b_1, b_2 \geq 0$. We introduce non-negative slack variables $\zeta_i, \zeta_i^*, i = 1, 2, \dots, l$, and then normalize S-LP:

$$\begin{aligned}
 \text{S-LP:} \quad & \underset{\alpha_i^*, \alpha_i, \xi_i, \xi_i^*, b_1, b_2, \zeta_i, \zeta_i^*}{\text{Max}} \quad - \frac{1}{l} \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \frac{C}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) \\
 \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) (x_i \cdot x_j) - \xi_j + b_1 - b_2 + \zeta_j = \varepsilon + y_j \\
 & - \sum_{i=1}^l (\alpha_i - \alpha_i^*) (x_i \cdot x_j) - \xi_i^* - b_1 + b_2 + \zeta_j^* = \varepsilon - y_j \\
 & \alpha_i, \alpha_i^*, \xi_i, \xi_i^*, b_1, b_2, \zeta_i, \zeta_i^* \geq 0, i = 1, 2, \dots, l.
 \end{aligned}$$

Let $a_{ij} = (x_i, x_j)$, $A_{ll} = (a_{ij})_{ll}$, $d_1 = (\varepsilon + y_1, \varepsilon + y_2, \dots, \varepsilon + y_l)^T$, $d_2 = (\varepsilon - y_1, \varepsilon - y_2, \dots, \varepsilon - y_l)^T$, and $h = (1, \dots, 1)_{1l}$. Then the constraint condition for S-LP is:

$$\begin{pmatrix} A_{ll} & -A_{ll} & -E_{ll} & 0 & h^T & -h^T & E_{ll} & 0 \\ -A_{ll} & A_{ll} & 0 & -E_{ll} & -h^T & h^T & 0 & E_{ll} \end{pmatrix} \begin{pmatrix} \alpha \\ \alpha^* \\ \xi \\ \xi^* \\ b_1 \\ b_2 \\ \zeta \\ \zeta^* \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}. \quad (7)$$

Constraint condition (7) equals to the following condition:

$$\begin{pmatrix} A_{ll} & -A_{ll} & -E_{ll} & 0 & h^T & -h^T & E_{ll} & 0 \\ 0 & 0 & -E_{ll} & -E_{ll} & 0 & 0 & E_{ll} & E_{ll} \end{pmatrix} \begin{pmatrix} \alpha \\ \alpha^* \\ \xi \\ \xi^* \\ b_1 \\ b_2 \\ \zeta \\ \zeta^* \end{pmatrix} = \begin{pmatrix} d_1 \\ \beta \end{pmatrix}, \quad (8)$$

where $\beta = (2\varepsilon, \dots, 2\varepsilon)_{1l}^T$.

It is obvious that the coefficient vectors corresponding to α_i, α_i^* is linear dependent, thus it can not appear in the basis at the same time. Thus, there must be one 0 in α_i and α_i^* , which is consistent with the aforesaid conclusion. It is clear from the dual problem that when $\alpha_i \neq 0$ (resp., $\alpha_i^* \neq 0$), α_i (resp., α_i^*) must correspond to a support vector. Based on equivalency of the problem, the variable α_i (resp., α_i^*) in the aforesaid linear programming must be a basic variable, and its column vector must be in the optimal basis. However, it can not be inferred that the point corresponding to α_i or α_i^* must be a non-support vector point when $\alpha_i = 0$ and $\alpha_i^* = 0$, which is because if there is $\alpha_i = 0$ or $\alpha_i^* = 0$, α_i and α_i^* may still be in the basic variable.

To simplify the problem, we **suppose** that:

No support vectors appears in $\alpha_i = \alpha_i^ = 0$ relating to aforesaid problems. (A stronger assumption is that the S-LP model is non-degenerate.)*

Thus, the following conditions are equivalent:

- (1) (x_i, y_i) is a support vector;
- (2) α_i or α_i^* is the basic variable of the optimal solution;
- (3) $\alpha_i \neq 0$ or $\alpha_i^* \neq 0$;
- (4) The coefficient column vector corresponding to α_i or α_i^* is in the optimal basis.

There are two categories of support vectors. When $\alpha_i \neq C$ and $\alpha_i^* \neq C$, the support vector is on the boundary; while $\alpha_i = C$ or $\alpha_i^* = C$, the support vector is beyond the boundary, which is called the generalized support vector.

3. Data Perturbation and Optimality Condition

The changes in original data considered in this paper involve two aspects: (1) the decrease (missing) of original data; (2) the addition of certain data. Aforesaid two types of data perturbation influence the S-LP model in the following ways:

(1) When data point (x_i, y_i) is deleted, according to constraint condition (7), the i -th row and the $(i + l)$ -th row constraint equations of the S-LP model are removed, and the decreased variables are $\alpha_i, \alpha_i^*, \xi_i, \xi_i^*, \zeta_i, \zeta_i^*$.

(2) When data point (x_{l+1}, y_{l+1}) is added, according to constraint condition (8), there are two additional constraint conditions for S-LP and the added variables are $\alpha_{l+1}, \alpha_{l+1}^*, \xi_{l+1}, \xi_{l+1}^*, \zeta_{l+1}, \zeta_{l+1}^*$.

Sensitivity analysis mainly involves two aspects: firstly, how will the optimal solution of said problem change when one or more parameters of the model change? Secondly, what is the range of change of said parameters when the optimal solution or optimal basis of the problem is to be unchanged? This paper focuses on the condition for remaining unchanged of the support vectors, i.e., what is the range of change of said parameters when the optimal solution or optimal basis of the problem is to be unchanged (when the support vectors maintain unchanged).

We propose the following solution for general linear programming:

$$\begin{aligned} \text{LP1:} \quad & \text{Max} \quad Z = CX \\ & \text{s.t.} \quad A_{mn}X \leq b \\ & \quad \quad X \geq 0, \end{aligned}$$

where $X = (x_1, \dots, x_n)^T$, $C = (c_1, \dots, c_n)$, $b = (b_1, \dots, b_m)^T$, $A = (a_{ij})_{mn}$.

The dual problem is:

$$\begin{aligned} \text{Min} \quad & w = Yb \\ \text{s.t.} \quad & YA_{mn} \geq C \\ & Y \geq 0. \end{aligned}$$

We then introduce the slack variables $(x_{n+1}, \dots, x_{n+m})$. The consumption coefficient matrix has the augmented matrix of $(A, E) = (P_1, P_2, \dots, P_{m+n})$, for which the basis is B , $B^{-1}(A, E) = (d_{ij})_{m \times (n+m)}$, $B^{-1} = (b_{ij})_{nm}$.

Where the following conditions are met, 1) the optimality condition, i.e., the test numbers $\sigma_j = c_j - C_B B^{-1} P_j \leq 0$, $j = 1, 2, \dots, n + m$;

(2) the feasibility condition, i.e., $B^{-1}b \geq 0$, the objective function realizes the optimal value $\bar{Z} = C_B B^{-1}b$ and the optimal solution $X_B = B^{-1}b$, $X_N = 0$, where X_B is corresponding to the basic variables and X_N is corresponding the non-basic variables. When LP1 is the optimal, the solution to the dual problem is $Y = (y_1, y_2, \dots, y_m) = C_B B^{-1}$, corresponding to the checking number of the slack variables. For more details on sensitivity analysis LP consumption coefficient matrix, the reader is referred to [5].

4. Addition of a Data Point (x_{l+1}, y_{l+1})

This section lays emphasis on the condition for remaining the support vectors unchanged when a data point (x_{l+1}, y_{l+1}) is added. For the programming S-LP, add the variables $\alpha_{l+1}, \alpha_{l+1}^*, \xi_{l+1}, \xi_{l+1}^*, \zeta_{l+1}, \zeta_{l+1}^*$, where $\zeta_{l+1}, \zeta_{l+1}^*$ are the slack variables; add the two constraint conditions $a_{l+1}\alpha - a_{l+1}\alpha^* - \xi_{l+1} - \xi_{l+1}^* + b_1 - b_2 + \zeta_{l+1} + \zeta_{l+1}^* = \varepsilon + y_{l+1}$ and $-\xi_{l+1} - \xi_{l+1}^* + \zeta_{l+1} + \zeta_{l+1}^* = 2\varepsilon$ in the $(2l+1)$ -th and $(2l+2)$ -th rows in (8), respectively.

The main conclusion of the section is stated as follows:

Theorem 1. *When added point (x_{l+1}, y_{l+1}) admits*

$$\begin{cases} -a^{l+1}\bar{X} + \varepsilon + y_{l+1} \geq 0, \\ a^{l+1}\bar{X} + 2\varepsilon \geq 0, \\ |\Delta| \leq \frac{1}{l}, \end{cases}$$

neither the support vectors nor the fitting function changes, where $\bar{X}$ is the optimal solution of the programming S-LP, and $\Delta = -C_B B^{-1} A_{l+1}$. Here, A_{l+1} is the column vector corresponding to α_{l+1} , while a^{l+1} is the coefficient vector of the $(2l+1)$ -th row in the constraints.

Before proving Theorem 1, we first apply LP1 to analyze the impact on the optimal basis of the addition of constraint conditions and the addition of variables. In which case, LP1 can be expressed as

$$\begin{aligned} \text{LP1:} \quad \text{Max} \quad Z &= (C_B, C_D) \begin{pmatrix} X_B \\ X_D \end{pmatrix} \\ \text{s.t.} \quad (B, D) \begin{pmatrix} X_B \\ X_D \end{pmatrix} &= b \\ X_B, X_D &\geq 0, \end{aligned}$$

where B is the optimal basis. According to the simplex method for linear programming, we imply

$$\begin{pmatrix} B & D & b \\ C_B & C_D & Z \end{pmatrix} \rightarrow \begin{pmatrix} I & B^{-1}D & B^{-1}b \\ 0 & C_D - C_B B^{-1}D & Z - C_B B^{-1}b \end{pmatrix}.$$

When $C_D - C_B B^{-1}D \leq 0$, $B^{-1}b \geq 0$, the optimal solution is $\bar{X} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}$.

(I) If one constraint condition $A_{m+1}X \leq b_{m+1}$ is added to LP1, where $A_{m+1} = (a_{m+1,1}, a_{m+1,2}, \dots, a_{m+1,n}) = ((A_{m+1})_B (A_{m+1})_D)$, we can introduce the slack variable x_{n+1} to convert the constraint to a equality constraint

$$(A_{m+1})_B X_B + (A_{m+1})_D X_D + X_{n+1} = b_{m+1}.$$

Thus, we obtain the new programming LP3.

$$\begin{aligned} \text{LP3: Max } Z &= (C_B \quad C_D \quad 0) \begin{pmatrix} X_B \\ X_D \\ X_{n+1} \end{pmatrix} \\ \text{s.t. } &\begin{pmatrix} B & D & 0 \\ (A_{m+1})_B & (A_{m+1})_D & 1 \end{pmatrix} \begin{pmatrix} X_B \\ X_D \\ X_{n+1} \end{pmatrix} = \begin{pmatrix} b \\ b_{m+1} \end{pmatrix} \\ &X_B, X_D, X_{n+1} \geq 0, \end{aligned}$$

where $B' = \begin{pmatrix} B & 0 \\ (A_{m+1})_B & 1 \end{pmatrix}$ is a new basis, $b' = \begin{pmatrix} b \\ b_{m+1} \end{pmatrix}$.

It implies $(B')^{-1} = \begin{pmatrix} B^{-1} & 0 \\ -(A_{m+1})_B B^{-1} & 1 \end{pmatrix}$ the checking number of

LP3 is

$$\begin{aligned} &(C_D \quad 0) - (C_B \quad 0)(B')^{-1} \begin{pmatrix} D \\ (A_{m+1})_D \end{pmatrix} \\ &= (C_D \quad 0) - (C_B \quad 0) \begin{pmatrix} B^{-1} & 0 \\ -(A_{m+1})_B B^{-1} & 1 \end{pmatrix} \begin{pmatrix} D \\ (A_{m+1})_D \end{pmatrix} \\ &= (C_D - C_B B^{-1} D \quad 0) \leq 0. \end{aligned}$$

Thus, in which case, the optimality condition is met.

Then, to see the feasibility: the basic solution is

$$\begin{aligned} \begin{pmatrix} X_B \\ X_{n+1} \end{pmatrix} &= (B')^{-1} \begin{pmatrix} b \\ b_{m+1} \end{pmatrix} = \begin{pmatrix} B^{-1} & 0 \\ -(A_{m+1})_B B^{-1} & 1 \end{pmatrix} \begin{pmatrix} b \\ b_{m+1} \end{pmatrix} \\ &= \begin{pmatrix} B^{-1}b \\ b_{m+1} - (A_{m+1})_B B^{-1}b \end{pmatrix}. \end{aligned}$$

As $B^{-1}b \geq 0$, thus when

$$b_{m+1} - (A_{m+1})_B B^{-1}b \geq 0, \quad (11)$$

we have $\begin{pmatrix} X_B \\ X_{n+1} \end{pmatrix} \geq 0$. In which case the condition for feasibility is met,

and $\begin{pmatrix} X_B \\ X_{n+1} \end{pmatrix}$ is the optimal solution.

(II) When a variable x_{n+1} is added to LP1, the latter can be rewritten as

$$\begin{aligned} \text{LP4: } \quad \text{Max} \quad Z &= (C_B \quad C_D \quad C_{n+1}) \begin{pmatrix} X_B \\ X_D \\ X_{n+1} \end{pmatrix} \\ \text{s.t.} \quad (C_B \quad C_D \quad C_{n+1}) &\begin{pmatrix} X_B \\ X_D \\ X_{n+1} \end{pmatrix} = b \\ X_B, X_D, X_{n+1} &\geq 0; \end{aligned}$$

$$\begin{pmatrix} B & D & D' & b \\ C_B & C_D & C_{n+1} & Z \end{pmatrix} \rightarrow \begin{pmatrix} I & B^{-1}D & B^{-1}D' & B^{-1}b \\ 0 & C_D - C_B B^{-1}D & C_{n+1} - C_B B^{-1}D' & Z - C_B B^{-1}b \end{pmatrix}.$$

Firstly, since $B^{-1}b \geq 0$, the condition for feasibility is met. Secondly, due to $C_D - C_B B^{-1}D \leq 0$, thus when

$$C_{n+1} - C_B B^{-1}D' \leq 0, \quad (12)$$

the condition for optimality is met. Hence both the optimal basis and the optimal solution are remained unchanged.

Let's prove Theorem 1 based on aforesaid analyses:

Proof. Let B be the optimal basis of S-LP and X_B be basic variable, and let $\bar{X}$ be the optimal solution, where $\bar{X}_B = B^{-1}D$ corresponding to B and $D = \begin{pmatrix} d_1 \\ \beta \end{pmatrix}$. Let the coefficient of the $(2l+1)$ -th row corresponding

to the original basic variable be a_B^{l+1} . It is obvious that the coefficient of the $(2l+2)$ -th row corresponding to the original basic variable is a zero vector. Thus, $X'_B = (X_B, \zeta_{l+1}, \zeta_{l+1}^*)$ is still a basic variable,

as its coefficient matrix $B' = \begin{pmatrix} B & 0 & 0 \\ a_B^{l+1} & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ is invertible and

$$B'^{-1} = \begin{pmatrix} B^{-1} & 0 & 0 \\ -a_B^{l+1}B^{-1} & 1 & 0 \\ a_B^{l+1}B^{-1} & 0 & 1 \end{pmatrix}, \text{ and it implies}$$

$$B'^{-1} \begin{pmatrix} D \\ \varepsilon + y_{l+1} \\ 2\varepsilon \end{pmatrix} = \begin{pmatrix} B^{-1}D \\ -a_B^{l+1}B^{-1}D + \varepsilon + y_{l+1} \\ a_B^{l+1}B^{-1}D + 2\varepsilon \end{pmatrix} = \begin{pmatrix} B^{-1}D \\ -a_B^{l+1}\bar{X}_B + \varepsilon + y_{l+1} \\ a_B^{l+1}\bar{X}_B + 2\varepsilon \end{pmatrix}.$$

As $B^{-1}D \geq 0$, thus when

$$\begin{cases} -a_B^{l+1}\bar{X} + \varepsilon + y_{l+1} \geq 0 \\ a_B^{l+1}\bar{X} + 2\varepsilon \geq 0 \end{cases} \quad (\Leftrightarrow) \quad \begin{cases} -a_B^{l+1}\bar{X}_B + \varepsilon + y_{l+1} \geq 0 \\ a_B^{l+1}\bar{X}_B + 2\varepsilon \geq 0 \end{cases},$$

the condition for feasibility is met, where a^{l+1} is the coefficient vector of row $2l+1$ in the constraints.

As to the optimality condition, condition (12) is sufficient according to the analysis on LP4, i.e., we need to inspect the checking numbers of α_{l+1} , α_{l+1}^* , ξ_{l+1} , ξ_{l+1}^* . The components of the vector $C_{l+1} = (-\frac{1}{l}, -\frac{1}{l}, -\frac{C}{l}, -\frac{C}{l})$ correspond to the checking numbers of variables α_{l+1} , α_{l+1}^* , ξ_{l+1} , ξ_{l+1}^* , respectively. Let the column vector corresponding to α_{l+1} as A_{l+1} , then the column vector corresponding to α_{l+1}^* is $-A_{l+1}$; ξ_{l+1} and

$$\xi_{l+1}^* \text{ correspond to the column vectors as } P_1 = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ -1 \end{pmatrix} \text{ and } P_2 = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ -1 \end{pmatrix},$$

respectively. Let $D' = (A_{l+1}, A_{l+1}, P_1, P_2)$ and $\Delta = -(C_B, 0, 0)B'^{-1}A_{l+1}$.

We imply $C_{l+1} - (C_B, 0, 0)B'^{-1}D' = (-\frac{1}{l} + \Delta, -\frac{1}{l} - \Delta, 0, 0)$.

Thus, when
$$\begin{cases} -\frac{1}{l} + \Delta \leq 0 \\ -\frac{1}{l} - \Delta \leq 0 \end{cases} \quad (\Leftrightarrow |\Delta| \leq \frac{1}{l}),$$
 the optimality condition is

met. In which case $B^{-1}D$ remains constant, thus the fitting function is unchanged. $\square$

5. Deletion of a Data Point (x_i, y_i)

When a point (x_i, y_i) is deleted from the original data, suppose that the point corresponds to a non-support vector. We need to find the conditions for not influencing the original support vectors. Then, delete the constraint conditions in the i -th and $(l+i)$ -th rows of S-LP. Delete six variables $\alpha_i, \alpha_i^*, \xi_i, \xi_i^*, \zeta_i, \zeta_i^*$. Since (x_i, y_i) corresponds to a non-support vector, according to the complementary slackness condition (4) given in the KKT condition of the SVR quadratic programming, this implies $\xi_i = \xi_i^* = 0$. Suppose S-LP is non-degenerate, then none of $\alpha_i, \alpha_i^*, \xi_i, \xi_i^*$ in S-LP are non-basic variables.

The main conclusion of this section is stated as follows:

Theorem 2. *Assume that the solution of S-LP is non-degenerate. If the solution of the dual problem of S-LP admits $h_i = h_{l+i} = 0$, then the optimal basis in the new programming with the non-support vector point (x_i, y_i) deleted is unchanged, the variables α_j, α_j^* in S-LP is still a basic variables, i.e., the support vectors remains unchanged. Moreover, since*

$B^{-1} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ is unchanged, the fitting function remains the same.

Analysis is first conducted on LP1. It is clear that to remove the i -th constraint condition $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$ from LP1, the equivalent method is to add a non-negative slack variable and a non-

negative residual variable to the i -th constraint condition, i.e.,
 $(+x_{n+1} - x_{n+2})$. Denoted the new programming by LP^* , then

$$Z = CX + 0x_{n+1} + 0x_{n+2},$$

$$LP^*: \quad \text{Max } Z = \begin{pmatrix} C_B & C_D & C_D' \end{pmatrix} \begin{pmatrix} X_B \\ X_D \\ X_D' \end{pmatrix}$$

$$\text{s.t.} \quad \begin{pmatrix} B & D & D' \end{pmatrix} \begin{pmatrix} X_B \\ X_D \\ X_D' \end{pmatrix} = b$$

$$X_B, X_D, X_D' \geq 0,$$

$$\text{where } C_D' = (0 \ 0), \ X_D' = \begin{pmatrix} X_{n+1} \\ X_{n+2} \end{pmatrix}, \ D' = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 1 & -1 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}.$$

With the simplex method, we implies

$$\begin{pmatrix} B & D & D' & b \\ C_B & C_D & C_D' & Z \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} I & B^{-1}D & B^{-1}D' & B^{-1}b \\ 0 & C_D - C_BB^{-1}D' & C_D' - C_BB^{-1}D' & Z - C_BB^{-1}b \end{pmatrix}.$$

It is obvious that the condition for feasibility is met. When $C_D^l - C_B B^{-1}D^l \leq 0$, the optimality condition is met, hence both the optimal basis and the optimal solution are unchanged.

Therefore, when a variable is deleted from LP1, the original basis remains the same and the conditions for feasibility and optimality are still met if the variable is not a basic variable.

Let's go back to the original problem. When a point (x_i, y_i) is deleted from the original data, constraints are removed from i -th and $(l+i)$ -th rows from S-LP and six variables $\alpha_i, \alpha_i^*, \xi_i, \xi_i^*, \zeta_i, \zeta_i^*$ are deleted as well. Based on aforesaid assumption, $\alpha_i, \alpha_i^*, \xi_i, \xi_i^*$ in S-LP are non-basic variables. Thus, we only need to discuss about the situation that ζ_i, ζ_i^* are in basis B . According to constraint condition (7) of S-LP, it is clear that the value coefficient vectors and coefficient vectors corresponding to

$$\xi_i, \xi_i^*, \zeta_i, \zeta_i^* \quad \text{are} \quad \left(-\frac{C}{l}, -\frac{C}{l}, 0, 0\right) \quad \text{and} \quad \begin{pmatrix} 0 \\ \vdots \\ -1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ -1 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix},$$

respectively.

Add non-negative slack variables $(+x_1^i - x_2^i)$ and $(+x_1^{l+i} - x_2^{l+i})$ separately to the i -th and $(l+i)$ -th rows, the coefficient vectors corresponding to $x_1^i, x_2^i, x_1^{l+i}, x_2^{l+i}$ are

$$\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ -1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ \vdots \\ -1 \\ \vdots \\ 0 \end{pmatrix}.$$

Let's prove Theorem 2 based on aforesaid analyses:

Proof. When a point (x_i, y_i) is deleted, there are two variables, α_i, α_i^* removed from S-LP, and the value coefficient of ξ_i, ξ_i^* is changed from $(-\frac{C}{l}, -\frac{C}{l})$ to $(0, 0)$. The basis B is unchanged (supplemented by the coefficient vector corresponding to the newly added slack variable), only that certain basic variables corresponding to B are replaced by newly added slack variables, thus the symbol of $\xi_i, \xi_i^*, \zeta_i, \zeta_i^*$ is still used as $x_1^i, x_2^i, x_1^{l+i}, x_2^{l+i}$. Let σ_1, σ_2 be the checking numbers of ξ_i, ξ_i^* before the perturbation. The solution of the dual problem is $H = (h_1, h_2, \dots, h_{2l})$.

Now let's discuss about the situation when ξ_i, ξ_i^* are in B :

(a) When both ζ_i and ζ_i^* are not in basis B , the basis is unchanged (after adding the coefficient vectors corresponding to the slack variables), while the feasibility is unchanged. The checking numbers of ζ_i, ζ_i^* are unchanged and separately equal to $-h_i, -h_{l+i}$. For the optimality, it is only needed to inspect the checking numbers σ'_1, σ'_2 of ξ_i, ξ_i^* . The solution is obtained as $\sigma'_1 = h_i, \sigma'_2 = h_{l+i}$. Thus the optimality is realized when $h_i = h_{l+i} = 0$.

(b) When only ζ_i (resp., ζ_i^*) is in the original basis B , the basis is unchanged; the feasibility is unchanged and the checking number $-h_{l+i}(-h_i)$ of $\zeta_i^*(\zeta_i)$ is unchanged. The checking numbers of ξ_i, ξ_i^* are $\sigma'_1 = h_i = 0, \sigma'_2 = h_{l+i}(\sigma'_1 = h_i, \sigma'_2 = h_{l+i} = 0)$.

Thus the optimality is realized when $h_{l+i} = 0(h_i = 0)$.

(c) When all ζ_i, ζ_i^* are in the original basis B , the checking numbers of ξ_i, ξ_i^* are $\sigma'_1 = h_i = 0, \sigma'_2 = h_{l+i} = 0$, which meets the optimality condition.

The theorem is proved thereby. $\square$

6. Conclusion

In this article, we first reduce the quadratic programming of SVR to linear programming. Based on sensitivity analysis of the linear programming, we then investigate data perturbation due to the decrease (missing) and increase of the initial data and provide the sufficient conditions for remaining the support vectors (the optimal basis) unchanged under the above cases of perturbation.

Acknowledgement

The research is supported by Beijing municipal colleges and universities high level talent introduction and training program (CIT&TCD201404080), and partially supported by scientific research management fund in Basic Courses Department of Beijing Union University (11202531501).

References

- [1] V. N. Vapnik, The Nature of Statistical Learning Theory (2nd Edition), Springer, New York, 2000.
- [2] Bernhard Schölkopf and Alexander J. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.
- [3] X. G. Zhang, Introduction to statistical learning theory and support vector machines, ACTA Automatica Sinica 26(1) (2000), 32-42.
- [4] Colin Campbell and Y. M. Ying, Learning with Support Vector Machines, Morgan and Claypool, 2011.
- [5] H. Li, Research on sensitivity analysis to LP consumption coefficient matrix, Mathematics in Practice and Theory 34(9) (2004), 119-126.