Journal of the Korean Data & Information Science Society
2014, **25**(2), 447–454

http://dx.doi.org/10.7465/jkdi.2014.25.2.447
한국데이터정보과학회지

Semi-supervised regression based on support vector machine[†]

Kyungha Seok[1]

[1]Department of Data Science, Inje University

Received 13 February 2014, revised 28 February 2014, accepted 6 March 2014

Abstract

In many practical machine learning and data mining applications, unlabeled training examples are readily available but labeled ones are fairly expensive to obtain. Therefore semi-supervised learning algorithms have attracted much attentions. However, previous research mainly focuses on classification problems. In this paper, a semi-supervised regression method based on support vector regression (SVR) formulation that is proposed. The estimator is easily obtained via the dual formulation of the optimization problem. The experimental results with simulated and real data suggest superior performance of the our proposed method compared with standard SVR.

Keywords: Semi-supervised regression, semi-supervised support vector regression, support vector regression, unlabeled data.

1. Introduction

One of the major problem in machine learning is how to get a large amount of labeled examples. However, data labeling is usually expensive due to the fact that labeling requires a lot of human efforts. While unlabeled data could yet relatively easy to obtain. For example, it is easy to download a batch of web pages from the internet, but it requires experts to label the pages. Therefore, semi-supervised learning (SSL), which employs both the labeled data and unlabeled data, has attracted a lot of research focus in recently year. One of the important issues in SSL is how to efficiently and effectively explore the information of the unlabeled data.

SSL have achieved successes in many applications, such as categorization, image classification, and spam filtering. The promising empirical success of SSL algorithms in favorable situations has triggered several recent attempts (Lafferty and Wasserman, 2007; Niyogi, 2008) at developing a theoretical understanding of SSL. In a recent paper, Singh *et al.* (2008) established that if the complexity of the distributions under consideration is too high to be understood using labeled data points, but is small enough to be understood using unlabeled data points, using a finite sample analysis in SSL can improve the performance of a supervised learning task. There have been many successful practical SSL algorithms

[†] This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0009705).

[1] Professor, Department of Data Science, Institute of Statistical Information, Inje University, Gyungnam 621-749, Korea. E-mail: statskh@inje.ac.kr

generated as summarized in Chapelle *et al.* (2006), Zhu (2005), Zhu and Goldberg (2009) and Seok (2010). It is worthwhile noting that the previous research focused primarily on classification and semi-supervised regression (SSR) remains largely under-studied.

According to the work of Belkin *et al.* (2006) the graph-based method could be applied to the regression estimator. This means that unlabeled data can contain helpful information and can increase the performance of the regression estimator. Zhou and Li (2007) proposed an SSR algorithm named COREG that boosts regression accuracy by exploiting unlabeled data in two k-nearest neighbor regression estimators using different distance metrics, each of which labels the unlabeled data for the other regression estimator. Wang *et al.* (2006) developed an SSR algorithm called semi-supervised kernel regression (SSKR) based on the classical kernel regression estimator. They also investigate the connection between the SSKR and the graph-based method. They showed that with a properly-chosen weighting factor, the SSKR remarkably outperformed kernel regression and the graph-based method. Cortes and Mohri (2007) dealt with regression problems in a transductive setting. They gave a new error boundary for transductive regression that holds for all bounded loss functions and coincides with the tight classification bounds of Vapnik (1998). Based on the given error bound, they presented a new algorithm for transductive regression that performs well and can scale to large data sets. Seok (2012, 2013) proposed semi-supervised local constant estimator (SSLCE) and kernel ridge estimator and showed the better performance of the semi-supervised method through numerical experiment. Seok (2012) also showed that the SSLCE has a faster convergence rate than that of the local constant estimator when a well chosen weighting factor is employed.

Some existing SSR methods have empirically shown promising performance. However, Lafferty and Wasserman (2008) showed that SSR methods based on regularization using graph Laplacians do not lead to faster minimax rates of convergence than those of kernel regression estimators. They also demonstrated how improved rates of convergence can be obtained by formulating and exploiting appropriate semi-supervised smoothness conditions.

Support vector machine (SVM) is being used as a powerful technique for regression and classification problems. 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 support vector regression (SVR) case, SRM results in a regularized ERM with an e-insensitive loss function. Introductions and applications of recent developments SVM can be found in Vapnik (1995, 1998), Smola and Schölkopf (1998), Shim and Hwang (2009) and Cristianini and Shawe-Taylor (2000). Xu *et al.* (2011) proposed a semi-supervised least squares SVR and showed their feasibility and efficiency by experiment on a Corn data set.

In this paper we derive a novel algorithm of semi-supervised learning for SVR based on SVM formulation. The estimator is easily obtained via the dual formulation of the optimization problem. In Section 2 we review SVR. In Section 3 we propose the semi-supervised SVR (S3VR) using SVR. In Section 4 we perform the numerical studies through synthetic and real examples. In Section 5 we give the conclusions.

2. Support vector regression

Let the training data set 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 output variable y_i is linearly or 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 regression function of the response given $\boldsymbol{x}_0$, $\mu(\boldsymbol{x}_0)$, can be regarded as a nonlinear function of input vector $\boldsymbol{x}_0$.

With e-insensitive loss function $\rho_e(\cdot)$, the estimator of the regression function can be defined as any solution to the optimization problem,

$$\min \frac{\lambda}{2}\boldsymbol{w}'\boldsymbol{w} + \sum_{i=1}^{n} \rho_e(y_i - \mu(\boldsymbol{x}_i)), \tag{2.1}$$

where $\lambda > 0$ is a regularization parameter, $\rho_e(r) = 0$ if $|r| \le e$ and $\rho_e(r) = |r| - e$ if $|r| > e$. We can express the regression problem (2.1) by formulation for SVM as follows:

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

subject to

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

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

$$\begin{aligned} L = \frac{1}{2}\boldsymbol{w}'\boldsymbol{w} + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}\alpha_i(e + \xi_i - y_i + \boldsymbol{w}'\phi(\boldsymbol{x}_i) + b) \\ - \sum_{i=1}^{n}\alpha_i^*(e + \xi_i^* + y_i - \boldsymbol{w}'\phi(\boldsymbol{x}_i) - b) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*). \end{aligned} \tag{2.2}$$

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 (2.2) with regard to the primal variables $(\boldsymbol{w}, \xi_i, \xi_i^*)$ and plugging them into equation (2.2), we have the optimization problem below.

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

with constraints $\alpha_i, \alpha_i^* \in [0, C]$, where the data points corresponding to positive values of α_i or α_i^* are called support vectors. Solving the above equation with the constraints determines

the optimal Lagrange multipliers, α_i, α_i^*, the estimator of the regression function given the input vector $\boldsymbol{x}_t$ is obtained as follows.

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

Here $\widehat{b}$ is obtained by KKT conditions (Kuhn and Tucker, 1951) as follows;

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

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

3. Semi-supervised support vector regression

Let the labeled training data set L denoted by $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$ and the unlabeled training data set U denoted by $\{\boldsymbol{x}_i\}_{i=n+1}^{n+m}$. Suppose the estimates of unlabeled responses were obtained in some way such as $\{\widetilde{y}_i\}_{i=n+1}^{n+m}$. We can express the semisupervised regression problem by formulation of SVM as follows:

$$\min \frac{\lambda}{2}\boldsymbol{w}'\boldsymbol{w} + C_1\sum_{i=1}^{n}(\xi_i + \xi_i^*) + C_2\sum_{i=1}^{m}(\zeta_i + \zeta_i^*)$$

subject to

$$\begin{aligned}
&y_i - \boldsymbol{w}'\phi(\boldsymbol{x}_i) - b \le e + \xi_i \\
&\boldsymbol{w}'\phi(\boldsymbol{x}_i) + b - y_i \le e + \xi_i^*, i = 1, \cdots, n \\
&\widetilde{y}_{n+i} - \boldsymbol{w}'\phi(\boldsymbol{x}_{n+i}) - b \le e + \zeta_i \\
&\boldsymbol{w}'\phi(\boldsymbol{x}_{n+i}) + b - \widetilde{y}_i \le e + \zeta_i^*, i = 1, \cdots, m,\ e, \xi_i^{(*)}, \zeta_i^{(*)} \ge 0
\end{aligned}$$

where C_1 and C_2 are penalty parameters penalizing the training errors. We construct a Lagrange function as follows:

$$\begin{aligned}
L = &\frac{\lambda}{2}\boldsymbol{w}'\boldsymbol{w} + \sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}\alpha_i(e + \xi_i - y_i + \boldsymbol{w}'\phi(\boldsymbol{x}_i) + b) \\
&- \sum_{i=1}^{n}\alpha_i^*(e + \xi_i^* + y_i - \boldsymbol{w}'\phi(\boldsymbol{x}_i) - b) - \sum_{i=1}^{n}\eta_i\xi_i - \sum_{i=1}^{n}\eta_i^*\xi_i^* \\
&- \sum_{i=1}^{m}\beta_i(e + \zeta_i - \widetilde{y}_i + \boldsymbol{w}'\phi(\boldsymbol{x}_{n+i}) + b) - \sum_{i=1}^{m}\beta_i^*(e + \zeta_i^* + \widetilde{y}_i - \boldsymbol{w}'\phi(\boldsymbol{x}_{n+i}) - b) \\
&- \sum_{i=1}^{m}\nu_i\zeta_i - \sum_{i=1}\nu_i^*\zeta_i^*
\end{aligned} \tag{3.1}$$

We notice that the positivity constraints $\alpha_i^{(*)}, \eta_i^{(*)}, \beta_i^{(*)}, \nu_i^{(*)} \geq 0$ should be satisfied. After taking partial derivatives of equation (3.1) with regard to the primal variables $(\boldsymbol{w}, \xi_i, \xi_i^*, \zeta_i, \zeta_i^*)$ and plugging them into equation (3.1), we have the optimization problem below.

$$\begin{aligned}\max & -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(\boldsymbol{x}_i,\boldsymbol{x}_j)-\frac{1}{2}\sum_{i,j=1}^{m}(\beta_i-\beta_i^*)(\beta_j-\beta_j^*)K(\boldsymbol{x}_{n+i},\boldsymbol{x}_{n+j})\\ & -\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i-\alpha_i^*)(\beta_j-\beta_j^*)K(\boldsymbol{x}_i,\boldsymbol{x}_{n+j})+\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)y_i+\sum_{i=1}^{m}(\beta_i-\beta_i^*)\widetilde{y_i}\\ & -e\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)-e\sum_{i=1}^{m}(\beta_i+\beta_i^*)\end{aligned}$$

with constraints $\alpha_i, \alpha_i^* \in [0, C]$, $\beta_i, \beta_i^* \in [0, C]$ and $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{m}(\beta_i - \beta_i^*) = 0$.

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

$$\widehat{\mu}(\boldsymbol{x}_t)=\sum_{i=1}^{n}(\widehat{\alpha}_i-\widehat{\alpha}_i^*)K(\boldsymbol{x}_i,\boldsymbol{x_t})+\sum_{i=1}^{m}(\widehat{\beta}_i-\widehat{\beta}_i^*)K(\boldsymbol{x}_{n+i},\boldsymbol{x_t})+\widehat{b}.$$

Here $\widehat{b}$ is obtained by KKT conditions (Kuhn and Tucker, 1951) as follows:

$$\begin{aligned}\widehat{b}=\frac{1}{n_{s_1}+n_{s_2}}\Big(&\sum_{i\in I_{s_1}}(y_i-K(\boldsymbol{x}_i,\boldsymbol{x}^l)(\alpha-\alpha^*)-K(\boldsymbol{x}_i,\boldsymbol{x}^u)(\beta-\beta^*))\\ &+\sum_{i\in I_{s_2}}(\widetilde{y_i}-K(\boldsymbol{x_{n+i}},\boldsymbol{x^l})(\widehat{\boldsymbol{\alpha}}-\widehat{\boldsymbol{\alpha^*}})-K(\boldsymbol{x_{n+i}},\boldsymbol{x^u})(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta^*}}))\Big)\end{aligned}$$

where n_{s_1} is the size of $I_{s_1} = \{i = 1, \cdots, n | 0 < \hat{\alpha}_i < C, 0 < \hat{\alpha}_i^* < C\}$, n_{s_2} is the size of $I_{s_2} = \{i = 1, \cdots, m | 0 < \widehat{\beta}_i < C, 0 < \widehat{\beta}_i^* < C\}$, $\boldsymbol{x}^l = (\boldsymbol{x}_i)_{i=1}^n$ and $\boldsymbol{x}^u = (\boldsymbol{x}_{n+i})_{i=1}^m$.

4. Numerical studies

We illustrate the performance of S3VR with SVR through the simulated and real data sets on the nonlinear regression cases. The radial basis kernel function is utilized in each example, which is,

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

In SVR $\widehat{\mu}(x_i)$ for $i = 1, \cdots, n$ is obtained by SVR with $\{y_i, x_i\}_{i=1}^n$, $\widehat{\mu}(x_i)$ for $i = n+1, \cdots, n+m$ is obtained by SVR with $\{y_i, x_i\}_{i=1}^n$ and $\{x_i\}_{i=n+1}^{n+m}$. In S3VR, $\widehat{\mu}(x_i)$ for $i = 1, \cdots, n+m$ is obtained by S3VR with $\{y_i, x_i\}_{i=1}^n$ and $\{\tilde{y}_i, x_i\}_{i=n+1}^{n+m}$, where $\widetilde{y_i} = \widehat{\mu}(x_i)$ is obtained by

SVR with $\{y_i, x_i\}_{i=1}^{n}$ and $\{x_i\}_{i=n+1}^{n+m}$. For simple calculation, we set $e = 0.1$ and the optimal values of (C, σ^2) for SVR and (C_1, C_2, σ^2) for S3VR are obtained from 10-fold cross validation function (Kohavi, 1995).

Example 4.1 To present the prediction performance of the proposed method 100 data sets are generated. Each data set consists of 100 x's and 100 y's. Here x's are generated from a uniform distribution U(0,1) and y's are generated from a normal distribution $N(1 + \sin(\pi x), 0.1)$. That is, the true regression function is given as

$$\mu(x) = 1 + \sin(\pi x).$$

Among 100 data, 80 unlabeled data ($m = 80$) are obtained by removing responses from a randomly chosen subset of 100 data, whereas the remaining 20 training data ($n = 20$) are treated as labeled. Figure 4.1 shows the true regression functions (solid lines), the estimated regression functions by SVR (dotted lines, left) and S3VR (dotted lines, right), respectively, superimposed on the scatter plots of one data set ($*$=labeled, o=unlabeled). From the Figure 4.1 we know that the unlabeled data could help to explore the relationship between x and y.

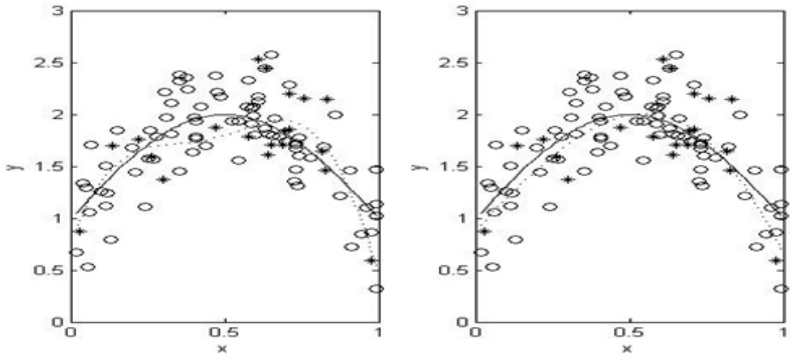

Figure 4.1 The regression functions superimposed on the scatter plots of 100 data points of a data set. The true regression functions, the estimated regression functions by SVR (left) and S3VR (right).

Example 4.2 For the second example 100 data sets are generated. Each data set consists of 100 x_1's, 100 x_2's and 100 y's. Here x_1's and x_2's are generated independently from a uniform distribution U(0,1) and y's are generated from a normal distribution $N(x_1 \exp(-0.5x_2), 0.1)$. The true regression function is given as

$$\mu(x_1, x_2) = x_1 \exp(-0.5x_2).$$

Among 100 data, 80 unlabeled data ($m = 80$) are obtained by removing responses from a randomly chosen subset of 100 data, whereas the remaining 20 data ($n = 20$) are treated as labeled.

With the data sets of Example 1 and Example 2 we obtain the mean squared error (MSE) to compare the performance of S3VR to SVR, which are shown in Table 4.1. The boldfaced figure signifies the smaller MSE in S3VR and SVR. From the table we can see that the proposed S3VR provides better performance than SVR.

Table 4.1 The average of 100 MSE of S3VR and SVR (standard deviations of MSE in parenthesis).

Example 1		Example 2	
SVR	S2SVR	SVR	S2SVR
0.0292 (0.0024)	**0.0264** (0.0023)	0.0846 (0.0077)	**0.0746** (0.0051)

Example 4.3 Corn data set (http://www.eigenvector.com/data/Corn/index.html) consists of 80 examples of corn measured on 3 different near-infra-red spectrometers, m5, mp5 and mp6. In this study, spectra from instrument m5 are used, where the wavelength range is 1100-2498nm at 2nm intervals. The moisture, oil, protein and starch values represent four output/dependent variables. As the first principal component describes 99% of the overall variance, this indicates high multicollinearity among the input/independent variables. Similar to Rosipal and Trejo (2001) and Xu *et al.* (2011), instead of modeling the real response we generated four different outputs as follows :

$$\begin{aligned} y_1 &= \exp(\boldsymbol{x}'\boldsymbol{x}/2m) \\ y_2 &= \exp(\boldsymbol{x}'A^{-\mathbf{1}}\boldsymbol{x}/2m1) \\ y_3 &= (\boldsymbol{x}'\boldsymbol{x}/m)^3 \exp(\boldsymbol{x}'\boldsymbol{x}/2m) \\ y_4 &= 0.3y_1 + 0.25y_2 - 0.7y_3 \end{aligned}$$

where A is a symmetric matrix with off-diagonal elements set to 0.8 and diagonal elements set to 1.0, and c and c_1 are averages of $\{\boldsymbol{x}_i'\boldsymbol{x}\}_{i=1}^{80}$ and $\{\boldsymbol{x}_i'A^{-1}\boldsymbol{x}\}_{i=1}^{80}$. The first 20 examples are used to create a training data set L, the last 20 examples are utilized to create a testing data set, and the remaining examples form a training data set U. In order to make the synthetic outputs (4.1) more realistic, Gaussian white noise with different levels is added, where noise level, denoted as n/s, corresponds to ratios of the standard deviation of the noise and the clean output variables. In this study, we set $n/s = 15\%$.

In order to assess prediction performance, we calculate the MSE and correlation coefficients (R) of actual and predicted outputs. Table 4.2 shows the result of Corn data experiment. The MSE and R of S3VR are less than those of SVR for all responses except y_3. From the Table 4.2, we see that our S3VR outperforms SVR.

Table 4.2 The average of MSE and R of S3VR and SVR for Corn data (standard deviations of MSE in parenthesis).

	MSE		R	
response	S3VR	SVR	S3VR	SVR
y_1	**0.1992** (0.0410)	0.2280 (0.0425)	**0.9882** (0.0079)	0.9845 (0.0083)
y_2	**0.3616** (0.0712)	0.4494 (0.0748)	**0.9649** (0.0239)	0.9575 (0.0255)
y_3	0.2505 (0.0559)	**0.2479** (0.0665)	0.9302 (0.0203)	**0.9335** (0.0228)
y_4	**0.4017** (0.2357)	0.4788 (0.2682)	**0.9300** (0.0707)	0.9141 (0.0822)

5. Concluding remarks

An interested algorithm of semi-supervised learning for SVR based on SVM formulation is proposed. In the proposed S3VR, the idea of kernel function is used to perform operations in the input space rather than in the high dimensional feature space. This enables us to handle

nonlinear as well as linear regression function estimation. We performed the numerical studies through synthetic and real data sets. The experiment revealed the superior performance of the proposed S3VR.

References

Belkin, M., Niyogi, P. and Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from laveled and unlabeled examples. *Journal of Machine Learning Research*, **1**, 1-48.

Chapelle, O., Schölkopf, B. and Zien, A. (2006). *Semi-supervised learning*, MIT Press, Cambridge, MA.

Cortes, C. and Mohri, M. (2007). On transductive regression. In *Advances in Neural Information Processing System*, **19**, 305-312.

Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines*, Cambridge University Press, United Kingdom.

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

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, **2**, Morgan Kaufmann, San Mateo, CA, 1137-1143.

Lafferty, J. and Wasserman, L. (2008). Statistical analysis of semi-supervised regression. In *Advances in Neural Information Processing Systems*, **20**, 801-808.

Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London A*, **209**, 415-446.

Niyogi, P. (2008). *Manifold regularization and semi-supervised learning: Some theoretical analyses*, Technical Report TR-2008-01, Computer Science Department, University of Chicago, Chicago, IL.

Rosipal, R. and Trejo, L. J. (2001). Kernel partial least squares regression in reproducing kernel Hilbert space. *Journal of Machine Learning Research*, **2**, 97-123

Seok, K. (2010). Semi-supervised classification with LS-SVM formulation. *Journal of Korean Data & Information Science Society*, **21**, 461-470.

Seok, K. (2012). Study on semi-supervised local constant regression estimation. *Journal of the Korean Data & Information Science Society*, **23**, 579-585.

Seok, K. (2013). A study on semi-supervised kernel ridge regression estimation. *Journal of the Korean Data & Information Science Society*, **24**, 341-353.

Shim, J. and Hwang, C. (2009). Support vector censored quantile regression under random censoring. *Computational Statistics and Data Analysis*, **53**, 912-919.

Singh, A., Nowak, R. and Zhu, X. (2008). Unlabeled data: Now it helps, now it doesn't. In *Advances in Neural Information Processing Systems*, **21**, 1513-1520.

Smola, A. and Schölkopf, B. (1998). On a kernel-based method for pattern recognition, regression, approximation and operator inversion. *Algorithmica*, **22**, 211-231.

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

Vapnik, V. N. (1998). *Statistical learning theory*, Wiley, New York.

Wang, M., Hua, X., Song, Y., Dai, L. and Zhang, H. (2006). Semi-supervised kernel regression. In *Proceeding of the Sixth IEEE International Conference on Data Mining*, 1130-1135.

Xu, S., An. X., Qiao, X., Zhu, L. and Li, L. (2011). Semisupervised least squares support vector regression machines. *Journal of Information & Computational Science*, **8**, 885-892.

Zhou, Z. and Li, M. (2007). Semi-supervised regression with co-training style algorithm. *IEEE Transactions on Knowledge and Data Engineering*, **19**, 1479-1493.

Zhu, D. (2005). *Semi-supervised learning literature survey*, Technical Report, Computer Sciences Department, University of Wisconsin, Madison, WI.

Zhu, X. and Goldberg, A. (2009). *Introduction to semi-supervised learning*, Morgan & Claypool, London.