The Eighth International Symposium on Operations Research and Its Applications (ISORA'09)
Zhangjiajie, China, September 20–22, 2009

On Universum-Support Vector Machines*

Ting-Ting Gao[1] Zhi-Xia Yang[2,3] Ling Jing[1,†]

[1]College of Science, China Agricultural University, 100083, Beijing, China
[2]College of Mathematics and System Science, Xinjiang University, Urumuchi 830046
[3]Academy of Mathematics and Systems Science, CAS, Beijing 100190

Abstract Universum-support vector machine (U-SVM) is an elegant method for 2-class classification problem. It is systematically studied in this paper, including the existence and uniqueness of the primal problem as well as the relation between the solutions of primal problem and dual problem. We find that U-SVM uses 3-class classification approach to solve the 2-class classification problem. So we have compared K-support vector classification regression (K-SVCR) and support vector ordinal regression machine (SVORM). Our conclusion is that, selecting their parameters properly, these three models get the same decision function essentially.

Keywords U-support vector machine; K-support vector classification regression; Support vector ordinal regression machine

1 Introduction

In this paper, we study a pattern classification algorithm which has recently been proposed in [1], called as the universum-support vector machine (U-SVM). So far, U-SVM has been successfully applied in many fields, such as handwriting digits recognition[1], gene translation initiation site identification[2] and so on. However, it is imperfect for the optimization theory of U-SVM. In this paper, we study its primal problem, dual problem and their relationship. In addition, for U-SVM, we delve into the relationship with both the K-support vector classification regression (K-SVCR)[6] and the support vector ordinal regression machine (SVORM)[7].

The rest of the paper is organized as follows. In Section 2, for U-SVM the primal problem, dual problem and their relationship are studied in detail . Of particular importance, we find the relationship among U-SVM, K-SVCR and SVORM, in Section 3. Section 4 contains some conclusion remarks.

2 U-SVM and its optimization theory

Suppose that the training set $\tilde{T}$ consists of two parts:

$$\tilde{T} = T_u \cup U_u, \tag{1}$$

*This work is supported by the Key Project of the National Natural Science Foundation of China (No.10631070), the National Natural Science Foundation of China (No.10801112) and the Ph.D Graduate Start Research Foundation of Xinjiang University Funded Project (No.BS080101).

†Corresponding author. E-mail: jingling@cau.edu.cn

where

$$T_u = \{(x_1,y_1),\cdots,(x_l,y_l)\} \in (\mathscr{X}\times\mathscr{Y})^l, \quad (2)$$
$$U_u = \{x_1^*,\cdots,x_u^*\} \in R^n, \quad (3)$$

with $x_i \in \mathscr{X} \subseteq R^n, y \in \mathscr{Y} = \{-1,1\}, i = 1,\cdots,l$, and $x_j^* \in R^n$, $j = 1,\cdots,u$. The task is to find a decision function

$$f(x) = \mathrm{sgn}((\tilde{w}\cdot x)+\tilde{b}). \quad (4)$$

2.1 The primal problem

The decision problem (4) is determined by $(\tilde{w},\tilde{b})$, a part of the solution $(\tilde{w},\tilde{b},\tilde{\xi},\tilde{\psi}^*)$ to the primal problem

$$\min_{w,b,\xi,\psi^{(*)}} \quad \frac{1}{2}\|w\|_2^2 + C_t\sum_{i=1}^{l}\xi_i + C_u\sum_{s=1}^{u}(\psi_s+\psi_s^*), \quad (5)$$
$$\text{s.t.} \quad y_i((w\cdot x_i)+b) \geq 1-\xi_i, \xi_i \geq 0, i = 1,\cdots,l, \quad (6)$$
$$-\varepsilon-\psi_s^* \leq (w\cdot x_s^*)+b \leq \varepsilon+\psi_s, s = 1,\cdots,u, \quad (7)$$
$$\psi_s,\psi_s^* \geq 0, s = 1,\cdots,u, \quad (8)$$

where $\xi = (\xi_1,\cdots,\xi_l)^T$, $\psi^{(*)} = (\psi_1,\psi_1^*,\cdots,\psi_u,\psi_u^*)^T$, and $C_t, C_u \in [0,+\infty), \varepsilon \in [0,+\infty)$ are the parameters. For the primal problem (5)–(8), we call $(\tilde{w},\tilde{b})$ above as the solution with respect to (w,b). Its existence and uniqueness are described below.

Theorem 1 For the primal problem (5)–(8), its solution $(\tilde{w},\tilde{b})$ with respect to (w,b) exists.

Theorem 2 For the primal problem (5)–(8), its solution $\tilde{w}$ with respect to w is unique , that is, if both $(w',b',\xi_i',\psi_i'^{(*)})$ and $(w'',b'',\xi_i'',\psi_i''^{(*)})$ are solutions of the primal problem (5)–(8), then $w'' = w'$.

Theorem 3 For the primal problem (5)–(8), its solutions with respect to b consist of a closed interval $[\tilde{b}^{dn},\tilde{b}^{up}]$.

2.2 The dual problem to the primal problem and their relationship

First, we give the dual problem of the primal problem (5)–(8).

Theorem 4 The optimization problem

$$\max_{\alpha,\mu,\nu} \quad -\frac{1}{2}\sum_{i,j=l}^{l}\alpha_i\alpha_j y_i y_j (x_i\cdot x_j) - \frac{1}{2}\sum_{s,t=1}^{u}(\mu_s-\nu_s)(\mu_t-\nu_t)(x_s^*\cdot x_t^*)$$
$$-\sum_{i=1}^{l}\sum_{s=1}^{u}\alpha_i y_i(\mu_s-\nu_s)(x_i\cdot x_s^*) + \sum_{i=1}^{l}\alpha_i - \varepsilon\sum_{s=1}^{u}(\mu_s+\nu_s) \quad (9)$$
$$\text{s.t.} \quad \sum_{i=1}^{l} y_i\alpha_i + \sum_{s=1}^{u}(\mu_s-\nu_s) = 0, \quad (10)$$
$$0 \leq \alpha_i \leq C_t, \ \ i = 1,\cdots,l, \quad (11)$$
$$0 \leq \mu_s,\nu_s \leq C_u, \ \ s = 1,\cdots,u\,. \quad (12)$$

is the dual problem of the primal problem (5)–(8).

In order to get the solution to primal problem, we need the following theorem.

Theorem 5 Suppose that $\tilde{\alpha} = (\tilde{\alpha}_1, \cdots, \tilde{\alpha}_l)^{\mathrm{T}}$, $\tilde{\mu} = (\tilde{\mu}_1, \cdots, \tilde{\mu}_u)^{\mathrm{T}}$, $\tilde{\nu} = (\tilde{\nu}_1, \cdots, \tilde{\nu}_u)^{\mathrm{T}}$ is any solution of the dual problem (9)–(12). If there exists $\tilde{\alpha}_j \in (0, C_t)$,or $\tilde{\mu}_m \in (0, C_u)$, or $\tilde{\nu}_t \in (0, C_u)$, then for the primal problem (5)–(8), its unique solution $(\tilde{w}, \tilde{b})$ with respect to (w, b) can be obtained by the following way:

$$\tilde{w} = \sum_{i=1}^{l} \tilde{\alpha}_i y_i x_i + \sum_{s=1}^{u} (\tilde{\mu}_s - \tilde{\nu}_s) x_s^*; \tag{13}$$

and

$$\tilde{b} = \begin{cases} y_j - \sum\limits_{i=1}^{l} \tilde{\alpha}_i y_i (x_i \cdot x_j) - \sum\limits_{s=1}^{u} (\tilde{\mu}_s - \tilde{\nu}_s)(x_s^* \cdot x_j), & \text{if } \tilde{\alpha}_j \in (0, C_t), \\ \varepsilon - \sum_{i=1}^{l} \tilde{\alpha}_i y_i (x_i \cdot x_m^*) - \sum_{s=1}^{u} (\tilde{\mu}_s - \tilde{\nu}_s)(x_s^* \cdot x_m^*), & \text{if } \tilde{\mu}_m \in (0, C_u), \\ -\varepsilon - \sum_{i=1}^{l} \tilde{\alpha}_i y_i (x_i \cdot x_t^*) - \sum_{s=1}^{u} (\tilde{\mu}_s - \tilde{\nu}_s)(x_s^* \cdot x_t^*), & \text{if } \tilde{\nu}_t \in (0, C_u). \end{cases} \tag{14}$$

3 The relationship with both K-SVCR and SVORM

Essentially speaking, U-SVM considers the U-set given by (3) as an extra class and deals with the 2-class problem by solving a 3-class problem. This leads to study their relationship. Here both K-SVCR and SVORM are addressed.

It has been pointed out in almost applications, only the case corresponding to Theorem 5 appears. So from now on, we assume that, for U-SVM, K-SVCR and SVORM, their solution with respect to (w, b) is unique.

3.1 The relationship between U-SVM and K-SVCR

Let us introduce K-SVCR[6] first, which was proposed to solve 3-class classification problem. Given a training set

$$\bar{T} = \{(x_i, y_i)\}_{i=1}^{l_1+l_2+l_3} \subseteq R^n \times \mathscr{Y}, \tag{15}$$

where $x_i \in R^n$, $y_i \in \mathscr{Y} = \{-1, 0, 1\}$ with

$$y_i = \begin{cases} -1, & i = 1, \cdots, l_1; \\ 0, & i = l_1 + 1, \cdots, l_1 + l_2; \\ 1, & i = l_1 + l_2 + 1, \cdots, l_1 + l_2 + l_3. \end{cases} \tag{16}$$

The task is to find a decision function

$$f(x) = \begin{cases} +1, & \text{if } (\bar{w} \cdot x) + \bar{b} \geq \delta; \\ -1, & \text{if } (\bar{w} \cdot x) + \bar{b} \leq -\delta; \\ 0, & \text{Others,} \end{cases}$$

where, $(\bar{w},\bar{b})$ is a solution with respect to (w,b) of the following problem with the parameters $D_1, D_2 \in [0,+\infty), \delta \in [0,1)$

$$\min_{w,b,\xi,\psi^{(*)}} \quad \frac{1}{2}\|w\|^2 + D_1 \sum_{i=1}^{l_1+l_2} \xi_i + D_2 \sum_{i=1}^{l_3} (\psi_i + \psi_i^*), \tag{17}$$

$$\text{s.t.} \quad y_i((w \cdot x_i)) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \cdots, l_1 + l_2, \tag{18}$$

$$-\delta - \psi_i^* \leq (w \cdot x_i^*)) + b \leq \delta + \psi_i, i = 1, \cdots, l_3, \tag{19}$$

$$\psi_i, \psi_i^* \geq 0, i = 1, \cdots, l_3, \tag{20}$$

where $\xi = (\xi_1, \cdots, \xi_{l_1}, \xi_{l_1+1}, \cdots, \xi_{l_1+l_2})^T, \psi^{(*)} = (\psi_1^{(*)}, \cdots, \psi_{l3}^{(*)})^T$.

Now we consider to use the above K-SVCR to solve the 2-class classification problem with the training set $\tilde{T}$ given by (2) and (3). First, transfer the training set $\tilde{T}$ into the formulation $\bar{T}$ given by (15) as following:

$$\bar{T} = \{(x_i, y_i)_{i=1}^{l_1+l_2+l_3}\}, \tag{21}$$

where

$$\{(x_i, y_i)_{i=1}^{l_1}\} = \{(x_i, y_i) | (x_i, y_i) \in T_u, y_i = -1\}, \tag{22}$$

$$\{(x_i, y_i)_{i=l_1+1}^{l_1+l^2}\} = \{(x_i, 0) | x_i \in U_u\}, \tag{23}$$

$$\{(x_i, y_i)_{i=l_1+l_2+1}^{l_1+l_2+l_3}\} = \{(x_i, y_i) | (x_i, y_i) \in T_u, y_i = 1\}. \tag{24}$$

Second, establish and solve the optimization problem (17)–(20) with the training set (21)–(24), and obtain its solution $(\bar{w},\bar{b})$ with respect to (w,b).

At last, construct the decision function

$$f(x) = \begin{cases} +1 & \text{when } (\bar{w} \cdot x) + \bar{b} \geq 0\ ; \\ -1 & \text{when } (\bar{w} \cdot x) + \bar{b} < 0\ , \end{cases} \tag{25}$$

Now we are in a position to show the relationship between the primal problem of U-SVM and the primal problem of K-SVCR:

Theorem 6 Suppose that, for the primal problem (5)–(8) and the primal problem (17)–(20) with the training set (21)–(24), $(\tilde{w},\tilde{b})$ and $(\bar{w},\bar{b})$ are respectively their solution with respect ro (w,b). If the parameters satisfy

$$C_t = D_1, C_u = D_2, \varepsilon = \delta,$$

then we have

$$\tilde{w} = \bar{w}, \ \ \tilde{b} = \bar{b}.$$

Therefore, the same decision function should be generated by U-SVM and the modified K-SVCR.

Proof. If $C_t = D_1, C_u = D_2, \varepsilon = \delta$, we can see that the primal problem (5)–(8) of U-SVM and the primal problem (17)–(20) of K-SVCR have the same form, so when the two models are trained for the training set (1), the same solutions are got, that is, $\tilde{w} = \bar{w}, \ \tilde{b} = \bar{b}$.

The decision function of U-SVM is :

$$(\tilde{w} \cdot x) + \tilde{b} = 0. \tag{26}$$

The decision function of K-SVCR is

$$(\bar{w} \cdot x) + \bar{b} = 0. \tag{27}$$

As can be seen, the both equations (26) and (27) represent the same hyperplane. □

3.2 The relationship between U-SVM and SVORM

Now we introduce SVORM solving a 3-class classification problem. Given a training set

$$\hat{T} = \{x_1^1, \cdots, x_{l^1}^1, x_1^2, \cdots, x_{l^2}^2, x_1^3, \cdots, x_{l^3}^3\}, \tag{28}$$

where $x_i^j \in R^n$ be the set of training examples, $j = 1, 2, 3$ denotes the class number, and l^j is the index within each class. The geometric interpretation of this approach is to look for 2 parallel hyperplanes

$$(\hat{w} \cdot x) = \hat{b}_s, s = 1, 2,$$

where $\hat{w} \in R^n$, $\hat{b}_1 \leq \hat{b}_2$, $\hat{b}_0 = -\infty$, and $\hat{b}_3 = +\infty$. So the space R^n is divided into 3 ranked regions by the decision rule

$$\hat{b}_{s-1} < (\hat{w} \cdot x) < \hat{b}_s, s = 1, 2, 3.$$

The decision function is then given by

$$f(x) = \min_{s \in \{1,2\}} \{s : (\hat{w} \cdot x) - \hat{b}_s < 0\}, \tag{29}$$

where $\hat{b}_0 = -\infty$, $\hat{b}_3 = +\infty$ and $(\hat{w}, \hat{b}_1, \hat{b}_2)$ is a solution with respect to (w, b_1, b_2) of the following primal problem with the parameters C_{13} and C_2

$$\min_{w, b_1, b_2, \xi^1, \xi^{(*)2}, \xi^{*3}} \quad \frac{1}{2}\|w\|^2 + C_{13}\left(\sum_{i=1}^{l^1} \xi_i^1 + \sum_{i=1}^{l^3} \xi_i^3\right) + C_2 \sum_{i=1}^{l^2} (\xi_i^2 + \xi_i^{*2}) \tag{30}$$

$$\text{s.t.} \quad ((w \cdot x_i^1) - b_1) \leq -1 + \xi_i^1, i = 1, \cdots, l^1, \tag{31}$$

$$((w \cdot x_i^2) - b_1) \geq 1 - \xi_i^{*2}, i = 1, \cdots, l^2, \tag{32}$$

$$((w \cdot x_i^2) - b_2) \leq -1 + \xi_i^2, i = 1, \cdots, l^2, \tag{33}$$

$$((w \cdot x_i^3) - b_2) \geq 1 - \xi_i^{*3}, i = 1, \cdots, l^3, \tag{34}$$

where $\xi^1 = (\xi_1^1, \cdots, \xi_{l^1}^1)^T, \xi^{(*)2} = (\xi_1^{(*)2}, \cdots, \xi_{l^2}^{(*)2})^T, \xi^3 = (\xi_1^{*3}, \cdots, \xi_{l^3}^{*3})^T$.

Now we consider to use the above SVORM to solve the 2-class classification problem with the training set $\hat{T}$ given by (2) and (3). First, transfer the training set $\tilde{T}$ into the formulation $\hat{T}$ given by (28) as following

$$\hat{T} = \{(x_i^j)_{i=1,\cdots,l_1+l_2+l_3}^{j=1,2,3}\}, \tag{35}$$

where

$$\{x_1^1,\cdots,x_{l_1}^1\} = \{x_i|(x_i,y_i)\in T_u, y_i=-1\}, \tag{36}$$

$$\{x_1^2,\cdots,x_{l^2}^2\} = \{x_i|x_i\in U_u\}, \tag{37}$$

$$\{x_1^3,\cdots,x_{l^3}^3\} = \{x_i|(x_i,y_i)\in T_u, y_i=1\}. \tag{38}$$

Second, establish and solve the optimization problem (30)–(34) with the training set (35)–(38), and obtain its solution $(\hat{w},\hat{b}_1,\hat{b}_2)$ with respect to (w,b_1,b_2).

At last, construct the decision function

$$f(x) = \begin{cases} +1, & \text{when } (\hat{w}\cdot x) - \frac{\hat{b}_1+\hat{b}_2}{2} \geq 0\ ; \\ -1, & \text{when } (\hat{w}\cdot x) - \frac{\hat{b}_1+\hat{b}_2}{2} < 0\ , \end{cases} \tag{39}$$

where, $(\hat{w},\hat{b_1},\hat{b_2})$ is the solution of (30)–(34).

Next, let us show the relationship between the primal problem of U-SVM and the primal problem of SVORM

Theorem 7 Suppose that, for the primal problem(5)–(8) and the primal problem (30)–(34) with the training set (35)–(38), $(\tilde{w},\tilde{b})$ and $(\hat{w},\hat{b}_1,\hat{b}_2)$ are respectively their solution with respect to (w,b) and (w,b_1,b_2). If the parameters satisfy

$$C_t = C_{13}, C_u = C_2, \varepsilon = \frac{\hat{b}_2-\hat{b}_1-2}{\hat{b}_2-\hat{b}_1+2},$$

then we have

$$\tilde{w} = \frac{\hat{w}}{\frac{2}{1-\varepsilon}}, \tilde{b} = \frac{-\frac{\hat{b}_1+\hat{b}_2}{2}}{\frac{2}{1-\varepsilon}}.$$

Therefore, the same decision function should be generated by U-SVM and the modified SVORM.

Proof. When 3-class classification SVORM is trained for the training set (35), the following both hyperplanes are no longer required

$$(w\cdot x) = b_1,\ (w\cdot x) = b_2,$$

but the hyperplane is located in the middle of above hyperplanes, that is,

$$(w\cdot x) - \frac{b_1+b_2}{2} = 0. \tag{40}$$

Then, the primal problem (30)–(34) of SVORM can be represented as

$$\min_{w,b_1,b_2} \quad \frac{1}{2}\|w\|^2 + C_{13}(\sum_{i=1}^{l^1}\xi_i^1 + \sum_{i=1}^{l^3}\xi_i^3) + C_2\sum_{i=1}^{l^2}(\xi_i^2+\xi_i^{*2}), \tag{41}$$

$$\text{s.t.} \quad (w\cdot x_i^1) - \frac{b_1+b_2}{2} \leq -(\frac{b_2-b_1}{2}+1) + \xi_i^1, i=1,\cdots,l^1, \tag{42}$$

$$(w\cdot x_i^2) - \frac{b_1+b_2}{2} \geq -(\frac{b_2-b_1}{2}-1) - \xi_i^{*2}, i=1,\cdots,l^2, \tag{43}$$

$$(w\cdot x_i^2) - \frac{b_1+b_2}{2} \leq (\frac{b_2-b_1}{2}-1) + \xi_i^2, i=1,\cdots,l^2, \tag{44}$$

$$(w\cdot x_i^3) - \frac{b_1+b_2}{2} \geq (\frac{b_2-b_1}{2}+1) - \xi_i^{*3}, i=1,\cdots,l^3. \tag{45}$$

Let $h = -\frac{b_1+b_2}{2}, \gamma = b_2 - b_1$. And given the proper normal direction $\widehat{w}$, the both margin hyperplanes can be represented as

$$(\widehat{w} \cdot x) + \widehat{b} = -(\frac{\gamma}{2} + 1) + \widehat{\xi}_i^1,$$
$$(\widehat{w} \cdot x) + \widehat{b} = (\frac{\gamma}{2} + 1) - \widehat{\xi}_i^{*3}.$$

For a fixed γ, let

$$w = \frac{\widehat{w}}{1+\frac{\gamma}{2}}, b = \frac{\widehat{b}}{1+\frac{\gamma}{2}}, \xi = \frac{\widehat{\xi}}{1+\frac{\gamma}{2}}.$$

Then, the optimization problem (41)–(45) is equivalent to the following problem

$$\min_{w,h,\xi} \quad \frac{1}{2}\|w\|^2 + C_{13}(\sum_{i=1}^{l^1} \xi_i^1 + \sum_{i=1}^{l^3} \xi_i^3) + C_2 \sum_{i=1}^{l^2} (\xi_i^2 + \xi_i^{*2}), \tag{46}$$

$$\text{s.t.} \quad (w \cdot x_i^1) + h \leq -1 + \xi_i^1, i = 1, \cdots, l^1, \tag{47}$$

$$-\frac{\gamma-2}{\gamma+2} - \xi_i^{*2} \leq (w \cdot x_i^2) + h \leq \frac{\gamma-2}{\gamma+2} + \xi_i^2, i = 1, \cdots, l^2, \tag{48}$$

$$(w \cdot x_i^3) + h \geq 1 - \xi_i^{*3}, i = 1, \cdots, l^3. \tag{49}$$

We can see that if $C_t = C_{13}, C_u = C_2, \varepsilon = \frac{\hat{b}_2 - \hat{b}_1 - 2}{\hat{b}_2 - \hat{b}_1 + 2}$, then the problem (46)–(49)is the same to the primal problem for U-SVM, and its solution with respect to (w,b) is $\tilde{w} = \frac{\hat{w}}{\frac{2}{1-\varepsilon}}, \tilde{b} = \frac{-\frac{\hat{b}_1+\hat{b}_2}{2}}{\frac{2}{1-\varepsilon}}$.

Furthermore, the decision function of U-SVM can be represented as

$$(\frac{\hat{w}}{\frac{2}{1-\varepsilon}} \cdot x) - \frac{\frac{\hat{b}_1+\hat{b}_2}{2}}{\frac{2}{1-\varepsilon}} = 0, \tag{50}$$

and the decision function of SVORM is

$$(\hat{w} \cdot x) - \frac{\hat{b}_1 + \hat{b}_2}{2} = 0. \tag{51}$$

we can see that the both equations (50) and (51) represent the same separating hyperplane. □

4 Discussion

In this paper we studies the theory of U-SVM, including the existence and uniqueness of the primal problem as well as the relation between the solutions of primal problem and dual problem. Of particular importance, we find that U-SVM is essentially use 3-class classifications to solve the 2-class classification problems, so we have U-SVM compared with K-SVCR and SVORM. Thus, we conclude that all of these three models get the same decision function when their parameters satisfy some conditions.

For simplicity, for U-SVM as well as K-SVCR and SVORM, only their linear formulations are addressed in this paper. However, it should be pointed out that for their nonlinear formulation with kernel, the same conclusions can be obtained without any essential difficulties.

References

[1] Weston, J., Collobert, R., Sinz,F., Bottou, L., and Vapnik, V. Inference with the universum, ICML, 2006.

[2] Gao, T.T., Tian, Y.J., Shao, X.J., and Deng, N.Y., Accurate Prediction of Translation Initiation Sites by Universum SVM, The Second International Symposium on Optimization and Systems Biology, 275–282, Lijiang, China, October 31– November 3, 2008.

[3] Vapnik, V. Estimation of dependences based on empirical data. Berlin:Springer Verlag. 2nd edition, 2006.

[4] Deng,N.Y.,and Tian,Y.J.,A New Method in Data Mining: Support Vector Machine. Science Press, Beijing (2004)

[5] Sinz, F.,Chapelle, O., Agarwal, A., and Schölkopf, B. An Analysis of Inference with the Universum. Advances in Neural Information Processing Systems 20: Proceedings of the 2007 Conference, 1369-1376, 2007.

[6] Angulo, C., and Parra, X. K-SVCR: A support vector machine for multi-class classifcation. Neurocomputing 55, 57-77, 2003.

[7] Shashua, A., and Levin, A. Taxonomy of large margin principle algorithms for ordinal regression problems. Technical Report, Leivniz Center for Research School of Computer Science and Eng., the hebrew University of Jerusaalem, 2002.

[8] Yang, Z.X., Support Vector Ordinal Regression Machines and Multi-class Classification. China Agricultural University, Ph.D Thesis, 2007.