
Vector-Valued Support Vector Regression (Paper #1751)

Mark James Brudnak, Ph.D.

U.S. Army Tank Automotive Research
Development and Engineering Center

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 18 JUL 2006		2. REPORT TYPE N/A		3. DATES COVERED	
4. TITLE AND SUBTITLE Vector-Valued Support Vector Regression				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Brudnak /Mark James, Ph.D.				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) USATACOM 6501 E 11 Mile Road Warren, MI 48397-5000				8. PERFORMING ORGANIZATION REPORT NUMBER 15999	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) TACOM TARDEC				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 34	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Vector-Valued Support Vector Regression

- The training data are of the form $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{\ell}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$
- $\hat{\mathbf{y}} = \mathbf{f}(\mathbf{x}, \boldsymbol{\pi})$, $\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m$
- Desire to minimize the loss functional (also called empirical risk functional)

$$\mathcal{J}(\boldsymbol{\pi}) = \sum_{i=1}^{\ell} L(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\pi}))$$

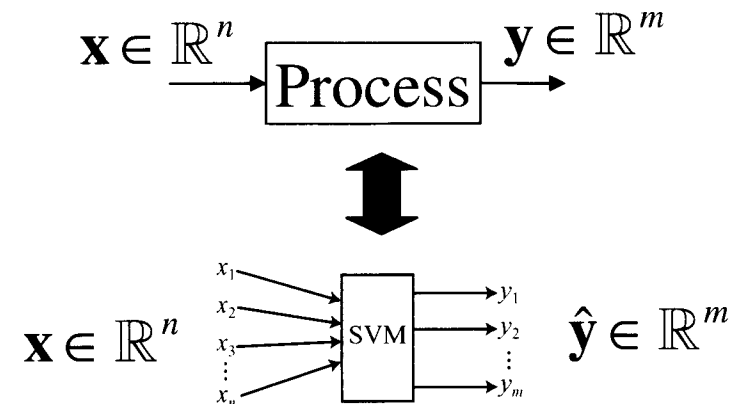
- Parameter preference expressed by regularization functional

$$\mathcal{P}(\boldsymbol{\pi})$$

- Balance the two as regularized risk functional

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C\mathcal{J}(\boldsymbol{\pi})$$

Handwritten note: $\mathcal{P}$



$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} L(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\pi}))$$

How should these be defined?

Parameters and estimator

Scalar Case: $y \in \mathbb{R}$

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + \sum_{i=1}^{\ell} L(y_i, \blacksquare(\mathbf{x}_i, \boldsymbol{\pi}))$$

Parameters:

$$\boldsymbol{\pi} = \{\mathbf{w}, b\}, \quad \mathbf{w} \in \mathbb{R}^v, b \in \mathbb{R}$$

- The weight $\mathbf{w}$ is free.
- The bias b is free.

Estimator form:



- The mapping $\boldsymbol{\varphi}(\cdot)$ is given.
- The estimator is linear in the range of $\boldsymbol{\varphi}(\cdot)$!

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + \sum_{i=1}^{\ell} L(\mathbf{y}_i, \blacksquare(\mathbf{x}_i, \boldsymbol{\pi}))$$

Parameters:

$$\boldsymbol{\pi} = \{\mathbf{W}, \mathbf{b}\}, \quad \mathbf{W} \in \mathbb{R}^{m \times v}, \mathbf{b} \in \mathbb{R}^m$$

- The weight $\mathbf{W}$ is free.
- The bias $\mathbf{b}$ is free.

Estimator form:



- The mapping $\boldsymbol{\varphi}(\cdot)$ is given.
- The estimator is linear in the range of $\boldsymbol{\varphi}(\cdot)$!

Regularization and Loss

Scalar Case: $y \in \mathbb{R}$

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} \mathbb{I}(y_i, f(\mathbf{x}_i, \boldsymbol{\pi}))$$

Regularization functional:

$$\mathcal{P}(\boldsymbol{\pi}) = \frac{1}{2} \mathbf{w}^T \mathbf{w}$$

- The weight $\mathbf{w}$ penalized.
- The bias b is not penalized.

Loss Function:



Vector Case: $\mathbf{y} \in \mathbb{R}^m$

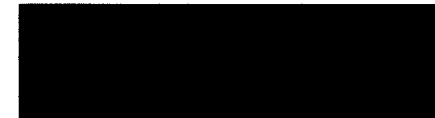
$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} \mathbb{I}(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\pi}))$$

Regularization functional:

$$\mathcal{P}(\boldsymbol{\pi}) = \frac{1}{2} \text{Tr}(\mathbf{W}\mathbf{W}^T)$$

- The weight $\mathbf{W}$ penalized.
- The bias $\mathbf{b}$ is not penalized.

Loss Function:



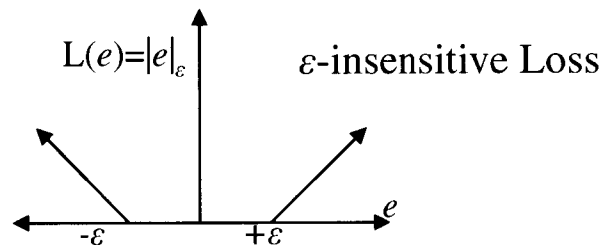
More to follow

The Loss Function (continued)

Scalar Case: $y \in \mathbb{R}$

$$e \triangleq y - \hat{y}$$

$$L(e) \triangleq |e|_\varepsilon = \max(0, |e| - \varepsilon)$$



Vector Case: $\mathbf{y}_i \in \mathbb{R}^m$

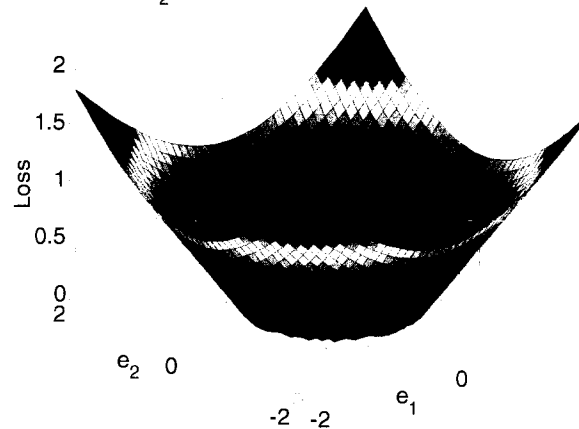
$$\mathbf{e} \triangleq \mathbf{y} - \hat{\mathbf{y}}$$

$$L(\mathbf{e}) \triangleq \|\mathbf{e}\|_p|_\varepsilon = \max(0, \|\mathbf{e}\|_p - \varepsilon)$$

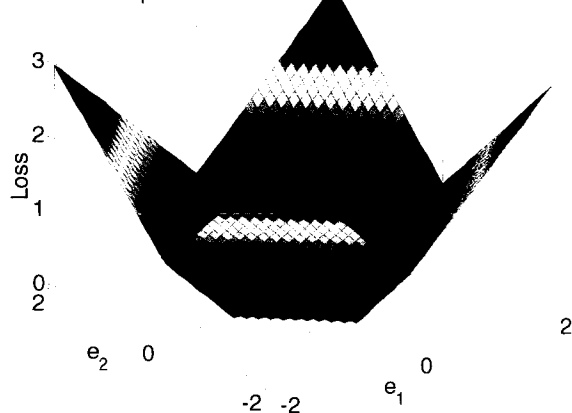
$$\|\mathbf{e}\|_p \triangleq \begin{cases} \left(\sum_{j=1}^m |e_j|^p \right)^{\frac{1}{p}} & 1 \leq p < \infty \\ \max_j (|e_j|) & p \sim \infty \end{cases}$$

ε_p -insensitive Loss

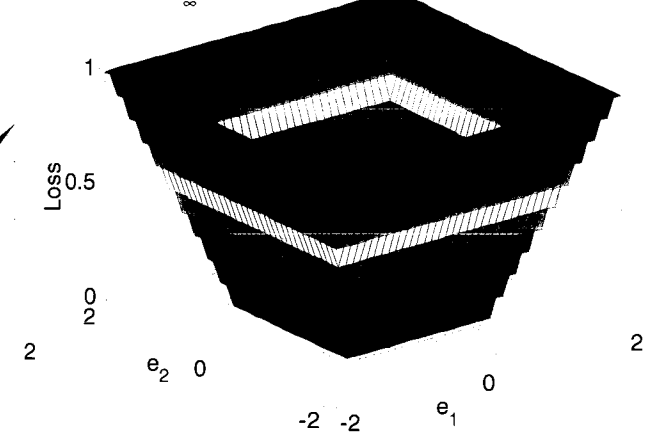
ε_2 -insensitive loss function with $\varepsilon = 1$



ε_1 -insensitive loss function with $\varepsilon = 1$



ε_∞ -insensitive loss function with $\varepsilon = 1$



Put it all Together

Scalar Case: $y \in \mathbb{R}$

Regularized Risk Functional:

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} L(y_i, f(\mathbf{x}_i, \boldsymbol{\pi}))$$



$$R_{reg} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{\ell} |y_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) - b|_{\varepsilon}$$

Problem:

Minimize: $R_{reg}(\mathbf{w}, b)$
 $\{\mathbf{w}, b\}$

Given: $\{\mathbf{x}_i, y_i\}_{i=1}^{\ell}, C, \varepsilon, \boldsymbol{\phi}(\cdot)$



Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Regularized Risk Functional:

$$R_{reg} = \mathcal{P}(\boldsymbol{\pi}) + C \sum_{i=1}^{\ell} L(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i, \boldsymbol{\pi}))$$



$$R_{reg} = \frac{1}{2} \text{Tr}(\mathbf{W}\mathbf{W}^T) + C \sum_{i=1}^{\ell} \|\mathbf{y}_i - \mathbf{W}\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{b}\|_p|_{\varepsilon}$$

Problem:

Minimize: $R_{reg}(\mathbf{W}, \mathbf{b})$
 $\{\mathbf{W}, \mathbf{b}\}$

Given: $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{\ell}, C, \varepsilon, \|\cdot\|_p, \boldsymbol{\phi}(\cdot)$



But what about the mapping $\boldsymbol{\phi}(\cdot)$?

Back to SVM: How to solve it

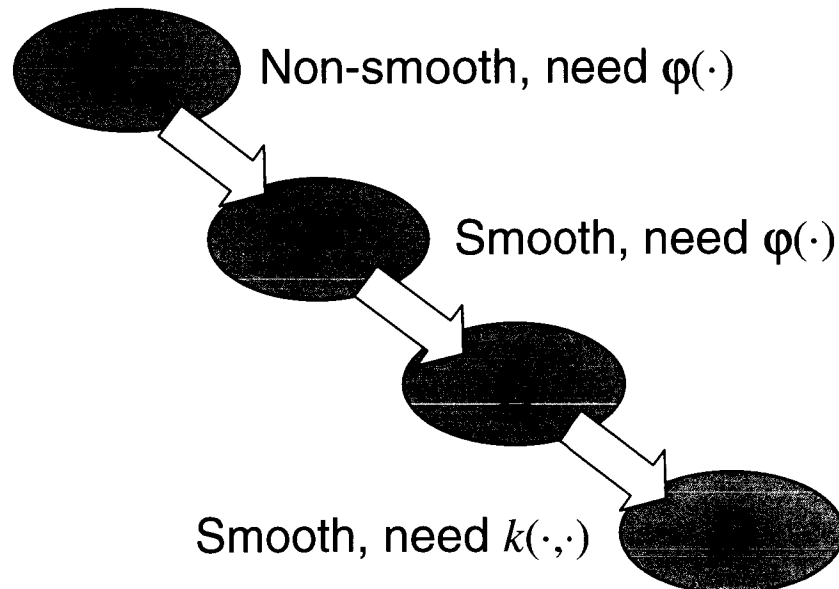
Scalar Case: $y \in \mathbb{R}$

Problem:

Minimize:
 $\{\mathbf{w}, b\}$

$$R_{reg} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{\ell} |y_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) - b|_{\varepsilon}$$

Approach:



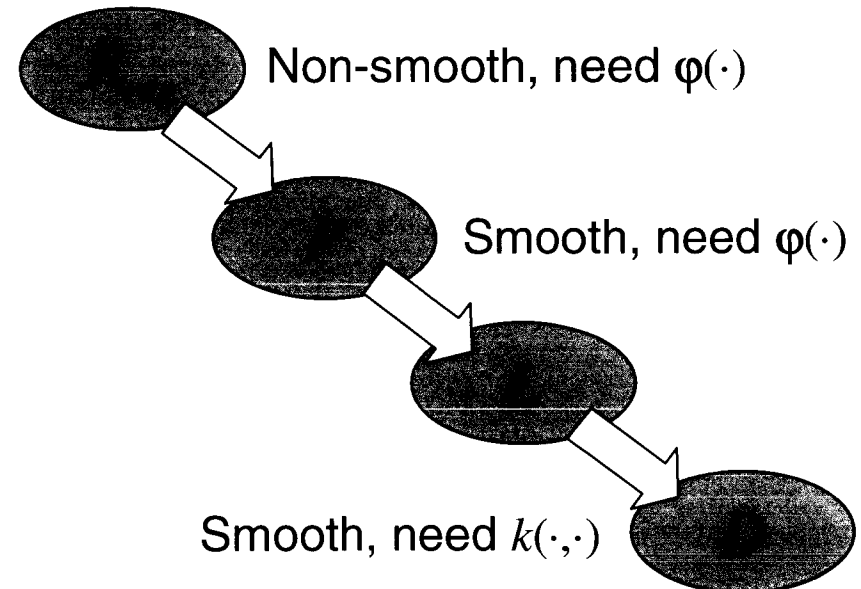
Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Problem:

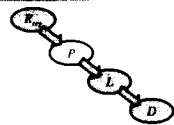
Minimize:
 $\{\mathbf{W}, \mathbf{b}\}$

$$R_{reg} = \frac{1}{2} \text{Tr}(\mathbf{W}\mathbf{W}^T) + C \sum_{i=1}^{\ell} \left\| \mathbf{y}_i - \mathbf{W}\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{b} \right\|_p \Big|_{\varepsilon}$$

Approach:



The Primal Problem



Scalar Case: $y \in \mathbb{R}$

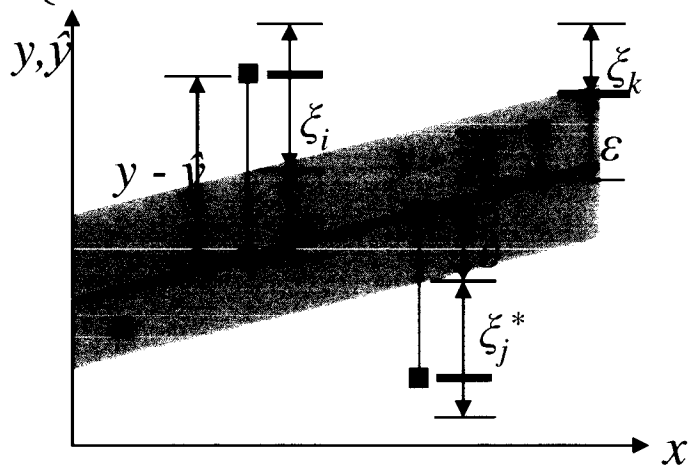
Minimize:

$$\mathbf{w}, b, \{\xi_i, \xi_i^*\}_{i=1}^{\ell}$$

$$P = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{\ell} (\xi_i + \xi_i^*)$$

subject to:

$$\begin{cases} y_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) - b - \varepsilon - \xi_i \leq 0 \\ -y_i + \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) + b - \varepsilon - \xi_i^* \leq 0 \\ \xi_i, \xi_i^* > 0 \end{cases}$$



Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Minimize:

$$\mathbf{W}, \mathbf{b}, \{\xi_i, \boldsymbol{\delta}_i, \boldsymbol{\delta}_i^*\}_{i=1}^{\ell}$$

Primal Variables

$$P = \frac{1}{2} \text{Tr}(\mathbf{W}\mathbf{W}^T) + C \sum_{i=1}^{\ell} \xi_i$$

subject to:

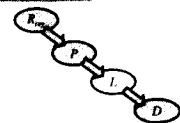
$$\begin{cases} \|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\|_p - \varepsilon - \xi_i \leq 0, & \xi_i \geq 0 \\ \mathbf{y}_i - \mathbf{W}\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{b} - \boldsymbol{\delta}_i \leq \mathbf{0}, & \boldsymbol{\delta}_i \geq \mathbf{0} \\ -\mathbf{y}_i + \mathbf{W}\boldsymbol{\phi}(\mathbf{x}_i) + \mathbf{b} - \boldsymbol{\delta}_i^* \leq \mathbf{0}, & \boldsymbol{\delta}_i^* \geq \mathbf{0} \end{cases}$$

$\boldsymbol{\delta}_i, \boldsymbol{\delta}_i^*$ and ξ_i are slack variables

Example: slack variable π

$$a \leq b \Leftrightarrow \begin{cases} a + \pi = b \\ \pi \geq 0 \end{cases}$$

The Lagrange Problem



Scalar Case: $y \in \mathbb{R}$

Minimize, Maximize:
 $\mathbf{w}, b, \{\xi_i, \xi_i^*\}_{i=1}^{\ell}$ $\{\alpha_i, \alpha_i^*, \eta_i, \eta_i^*\}_{i=1}^{\ell}$

$$L \triangleq \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{\ell} (\xi_i + \xi_i^*)$$

$$- \sum_{i=1}^{\ell} \alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b)$$

$$- \sum_{i=1}^{\ell} \alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b)$$

$$- \sum_{i=1}^{\ell} (\eta_i \xi_i + \eta_i^* \xi_i^*)$$

subject to: $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Minimize, Maximize:
 $\mathbf{w}, \mathbf{b}, \{\xi_i, \boldsymbol{\delta}_i, \boldsymbol{\delta}_i^*\}_{i=1}^{\ell}$ $\{\alpha_i, \eta_i, \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_i^*, \boldsymbol{\theta}_i, \boldsymbol{\theta}_i^*\}_{i=1}^{\ell}$

Dual
Variables

$$L \triangleq \frac{1}{2} \text{Tr}(\mathbf{W} \mathbf{W}^T) + C \sum_{i=1}^{\ell} \xi_i$$

$$+ \sum_{i=1}^{\ell} \alpha_i (\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\|_p - \varepsilon - \xi_i) - \sum_{i=1}^{\ell} \eta_i \xi_i$$

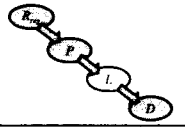
$$- \sum_{i=1}^{\ell} \boldsymbol{\gamma}_i^T (\mathbf{y}_i - \mathbf{W} \boldsymbol{\varphi}(\mathbf{x}_i) - \mathbf{b} - \boldsymbol{\delta}_i)$$

$$- \sum_{i=1}^{\ell} \boldsymbol{\gamma}_i^{*T} (-\mathbf{y}_i + \mathbf{W} \boldsymbol{\varphi}(\mathbf{x}_i) + \mathbf{b} - \boldsymbol{\delta}_i^*)$$

$$- \sum_{i=1}^{\ell} (\boldsymbol{\theta}_i^T \boldsymbol{\delta}_i + \boldsymbol{\theta}_i^{*T} \boldsymbol{\delta}_i^*)$$

subject to: $\alpha_i, \eta_i \geq 0, \quad \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_i^*, \boldsymbol{\theta}_i, \boldsymbol{\theta}_i^* \geq 0$

Minimization of L over Primals



Scalar Case: $y \in \mathbb{R}$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{\ell} (\alpha_i^* - \alpha_i) = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{\ell} (\alpha_i^* - \alpha_i) \mathbf{x}_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0$$

Introduce: $\beta_i = \alpha_i - \alpha_i^*$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$\frac{\partial L}{\partial \mathbf{b}} = \sum_{i=1}^{\ell} (\gamma_i^* - \gamma_i) = 0$$

$$\frac{\partial L}{\partial \mathbf{W}} = \mathbf{W} - \sum_{i=1}^{\ell} (\gamma_i - \gamma_i^*) \boldsymbol{\phi}^T(\mathbf{x}_i) = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0$$

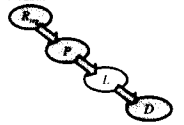
$$\frac{\partial L}{\partial \boldsymbol{\delta}_i} = \alpha_i \left(\frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*}{\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\|_p} \right)^{p-1} - \gamma_i - \boldsymbol{\theta}_i = 0$$

$$\frac{\partial L}{\partial \boldsymbol{\delta}_i^*} = \alpha_i \left(\frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*}{\|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\|_p} \right)^{p-1} - \gamma_i^* - \boldsymbol{\theta}_i^* = 0$$

Introduce: $\boldsymbol{\Gamma}_i = \gamma_i - \gamma_i^*$

$\mathbf{e}_i = \boldsymbol{\delta}_i - \boldsymbol{\delta}_i^*$

Implications of Optimization



Scalar Case: $y \in \mathbb{R}$

Equality Constraint:

$$\sum_{i=1}^{\ell} \beta_i = 0$$

Weights:

$$\mathbf{w} = \sum_{i=1}^{\ell} \beta_i \boldsymbol{\phi}(\mathbf{x}_i)$$

Constraints:

$$C - \alpha_i - \eta_i = 0 \rightarrow \alpha_i \leq C$$

$$C - \alpha_i^* - \eta_i^* = 0 \rightarrow \alpha_i^* \leq C$$

Estimator:

$$\hat{y} = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}) + b$$

$$= \sum_{i=1}^{\ell} \beta_i \boldsymbol{\phi}(\mathbf{x}_i)^T \boldsymbol{\phi}(\mathbf{x}) + b$$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Equality Constraint:

$$\sum_{i=1}^{\ell} \Gamma_i = 0$$

Weights:

$$\mathbf{W} = \sum_{i=1}^{\ell} \Gamma_i \boldsymbol{\phi}^T(\mathbf{x}_i)$$

Constraints:

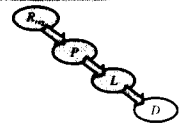
$$C - \alpha_i - \eta_i = 0 \rightarrow \alpha_i \leq C$$

Estimator:

$$\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{W} \boldsymbol{\phi}(\mathbf{x}) + \mathbf{b}$$

$$= \sum_{i=1}^{\ell} \Gamma_i \boldsymbol{\phi}(\mathbf{x}_i)^T \boldsymbol{\phi}(\mathbf{x}) + \mathbf{b}$$

The Dual Problem



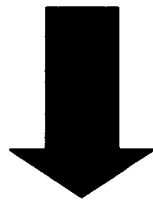
Scalar Case: $y \in \mathbb{R}$

Regularized Risk:

Minimize:
 $\{\mathbf{w}, b\}$

$$R_{reg} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{\ell} |y_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) - b|_{\varepsilon}$$

Dual Problem:



Maximize:
 $\{\beta_i\}_{i=1}^{\ell}$

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \beta_i \beta_j \mathbf{\Phi}(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} y_i \beta_i - \varepsilon \sum_{i=1}^{\ell} |\beta_i|$$

Subject to:

$$\sum_{i=1}^{\ell} \beta_i = 0, \quad |\beta_i| \leq C$$

Solve for β_i numerically

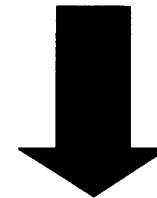
Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Regularized Risk:

Minimize:
 $\{\mathbf{W}, \mathbf{b}\}$

$$R_{reg} = \frac{1}{2} \text{Tr}(\mathbf{W}\mathbf{W}^T) + C \sum_{i=1}^{\ell} \|\mathbf{y}_i - \mathbf{W}\boldsymbol{\phi}(\mathbf{x}_i) - \mathbf{b}\|_p|_{\varepsilon}$$

Dual Problem:



Maximize:
 $\{\Gamma_i\}_{i=1}^{\ell}$

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \Gamma_i^T \Gamma_j \mathbf{\Phi}(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} \mathbf{y}_i^T \Gamma_i - \varepsilon \sum_{i=1}^{\ell} \|\Gamma_i\|_q$$

Subject to:

$$\sum_{i=1}^{\ell} \Gamma_i = 0, \quad \|\Gamma_i\|_q \leq C \quad \frac{1}{p} + \frac{1}{q} = 1$$

Solve for Γ_i numerically

Half Way There

Scalar Case: $y \in \mathbb{R}$

Dual Problem:

Solve for β_i
numerically

Maximize: $D(\{\beta_i\}_{i=1}^{\ell})$

Subject to: $\sum_{i=1}^{\ell} \beta_i = 0, \quad |\beta_i| \leq C$

Estimator:

$$\hat{y} = \sum_{i=1}^{\ell} \beta_i k(\mathbf{x}_i, \mathbf{x}) + b$$

Support Vectors: $\{\mathbf{x}_i : \beta_i \neq 0\}$

The bias remains to be found

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Dual Problem:

Solve for Γ_i
numerically

Maximize: $D(\{\Gamma_i\}_{i=1}^{\ell})$

Subject to: $\sum_{i=1}^{\ell} \Gamma_i = 0, \quad \|\Gamma_i\|_q \leq C$

Estimator:

$$\hat{\mathbf{y}}(\mathbf{x}) = \sum_{i=1}^{\ell} \Gamma_i k(\mathbf{x}_i, \mathbf{x}) + \mathbf{b}$$

Support Vectors: $\{\mathbf{x}_i : \Gamma_i \neq 0\}$

The bias remains to be found

Must Develop the KKT Conditions to find bias

KKT Conditions

Scalar Case: $y \in \mathbb{R}$

$$\begin{aligned}\alpha_i(\varepsilon + \xi_i - y_i + \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b) &= 0 \\ \alpha_i^*(\varepsilon + \xi_i^* + y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b) &= 0 \\ \eta_i \xi_i &= (C - \alpha_i) \xi_i = 0 \\ \eta_i^* \xi_i^* &= (C - \alpha_i^*) \xi_i^* = 0\end{aligned}$$

Note: $\circ$ denotes a parallel or element-wise product.
Exponents are taken element-wise

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$\begin{aligned}\alpha_i \left(\left\| \boldsymbol{\delta}_i + \boldsymbol{\delta}_i^* \right\|_p - \varepsilon - \xi_i \right) &= 0 \\ \eta_i \xi_i &= (C - \alpha_i) \xi_i = 0 \\ \boldsymbol{\gamma}_i \circ (\mathbf{y}_i - \mathbf{W} \boldsymbol{\varphi}(\mathbf{x}_i) - \mathbf{b} - \boldsymbol{\delta}_i) &= \mathbf{0} \\ \boldsymbol{\gamma}_i^* \circ (-\mathbf{y}_i + \mathbf{W} \boldsymbol{\varphi}(\mathbf{x}_i) + \mathbf{b} - \boldsymbol{\delta}_i^*) &= \mathbf{0} \\ \boldsymbol{\theta}_i \circ \boldsymbol{\delta}_i &= \left(\alpha_i \left(\frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*}{\left\| \boldsymbol{\delta}_i + \boldsymbol{\delta}_i^* \right\|_p} \right)^{p-1} - \boldsymbol{\gamma}_i^* \right) \circ \boldsymbol{\delta}_i = \mathbf{0} \\ \boldsymbol{\theta}_i^* \circ \boldsymbol{\delta}_i^* &= \left(\alpha_i \left(\frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*}{\left\| \boldsymbol{\delta}_i + \boldsymbol{\delta}_i^* \right\|_p} \right)^{p-1} - \boldsymbol{\gamma}_i^* \right) \circ \boldsymbol{\delta}_i^* = \mathbf{0}\end{aligned}$$

Implications of KKT Conditions

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Primal + Dual + KKT conditions imply

(absolute value and exponent are taken element wise)

$$\boldsymbol{\delta}_i \circ \boldsymbol{\delta}_i^* = \mathbf{0}$$

$$\boldsymbol{\gamma}_i \circ \boldsymbol{\gamma}_i^* = \mathbf{0}$$

$$\Gamma_{i,j} \neq 0 \rightarrow \text{sign}(\Gamma_{i,j}) = \text{sign}(e_{i,j})$$

$$|\boldsymbol{\Gamma}_i| = \alpha_i \frac{d}{d\mathbf{e}_i} (\|\mathbf{e}_i\|_p)$$

Lemma 1

$$\alpha_i = \|\boldsymbol{\Gamma}_i\|_q$$

$$\boldsymbol{\delta}_i - \boldsymbol{\delta}_i^* \equiv \mathbf{e}_i$$

$$\frac{1}{p} + \frac{1}{q} = 1$$

$$\frac{|\boldsymbol{\Gamma}_i|}{\|\boldsymbol{\Gamma}_i\|_q} = \left| \frac{\mathbf{e}_i}{\|\mathbf{e}_i\|_p} \right|^{p-1}, \quad \forall p \in [1, \infty]$$

or

$$\left| \frac{\boldsymbol{\Gamma}_i}{\|\boldsymbol{\Gamma}_i\|_q} \right|^q = \left| \frac{\mathbf{e}_i}{\|\mathbf{e}_i\|_p} \right|^p, \quad p \neq 1$$

KKT Conditions (in the tube)

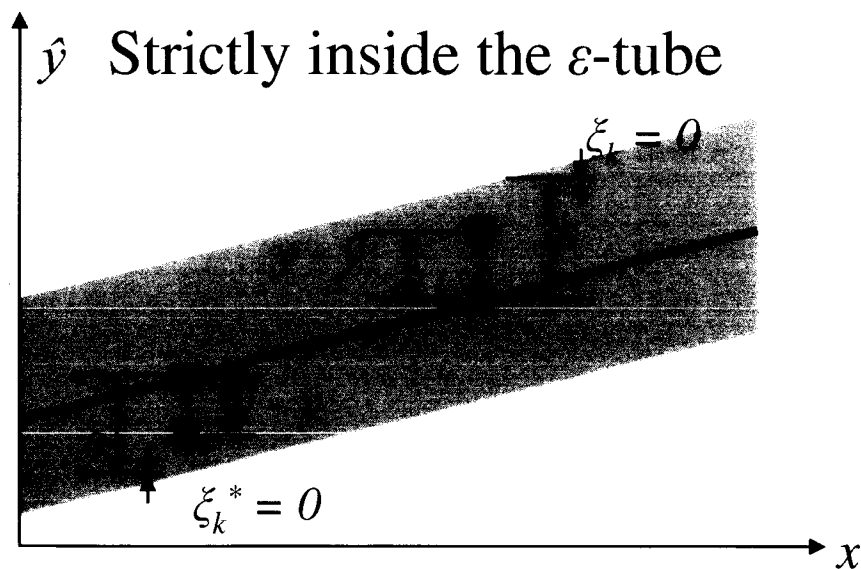
$$\frac{1}{p} + \frac{1}{q} = 1$$

Scalar Case: $y \in \mathbb{R}$

$$\alpha_i = \alpha_i^* = 0$$

$$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b = e_i < \varepsilon$$

$$\Rightarrow -y_i + \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b = -e_i < \varepsilon$$

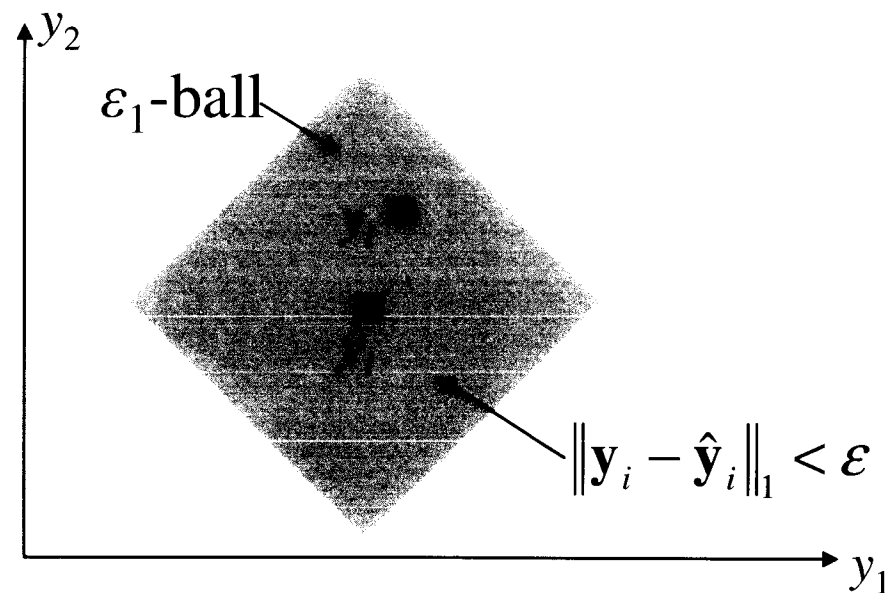


Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$\alpha_i = \|\boldsymbol{\Gamma}_i\|_q = 0$$

$$\Rightarrow \|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\|_p = \|\mathbf{e}_i\|_p < \varepsilon$$

Strictly inside the ε_p -ball



KKT Conditions (on the margin)

$$\frac{1}{p} + \frac{1}{q} = 1$$

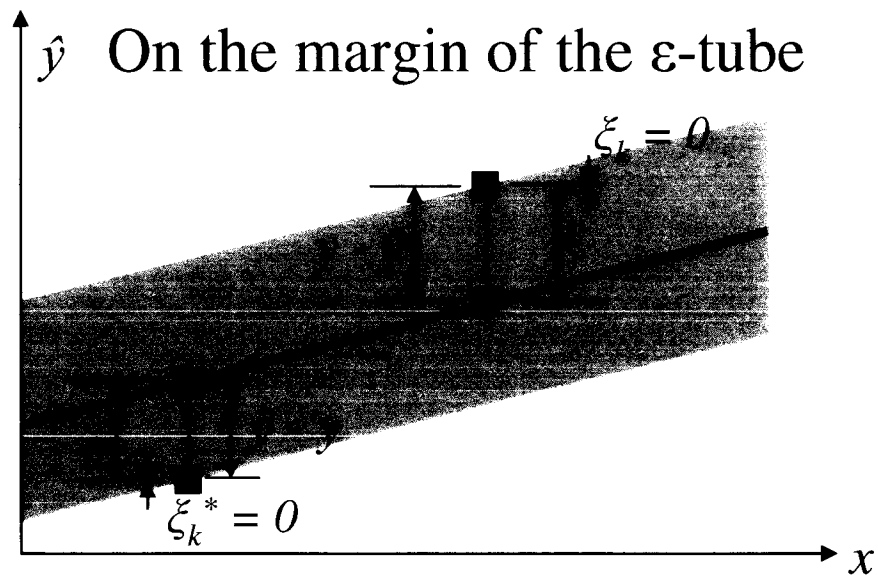
Scalar Case: $y \in \mathbb{R}$

$$\alpha_i \in (0, C)$$

$$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) - b \triangleq e_i = \varepsilon$$



$$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) - b \triangleq \text{[redacted]}$$

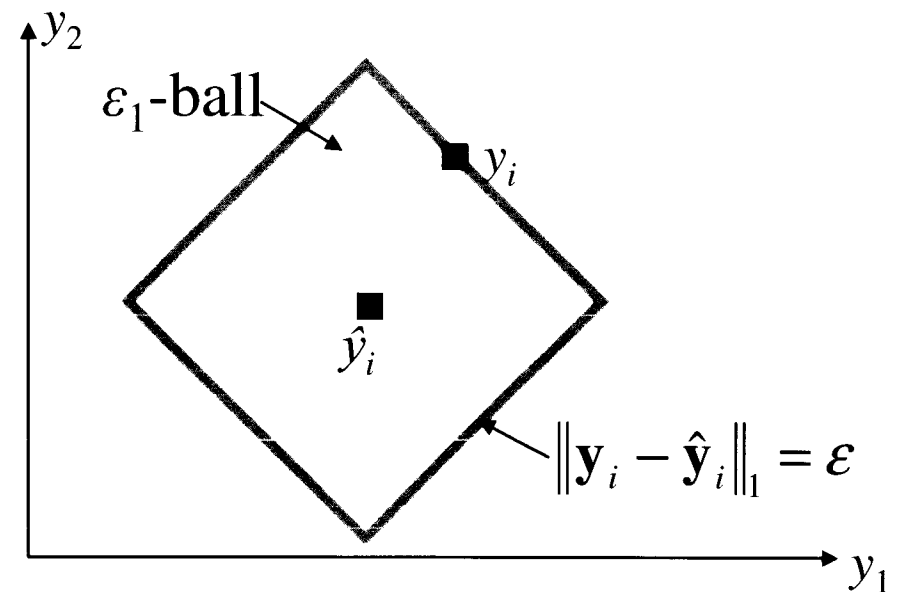


Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$\alpha_i = \|\boldsymbol{\Gamma}_i\|_q \in (0, C)$$

$$\Rightarrow \|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\|_p = \|\mathbf{e}_i\|_p = \varepsilon$$

On the boundary of the ε_p -ball



KKT Conditions (out of the tube)

$$\frac{1}{p} + \frac{1}{q} = 1$$

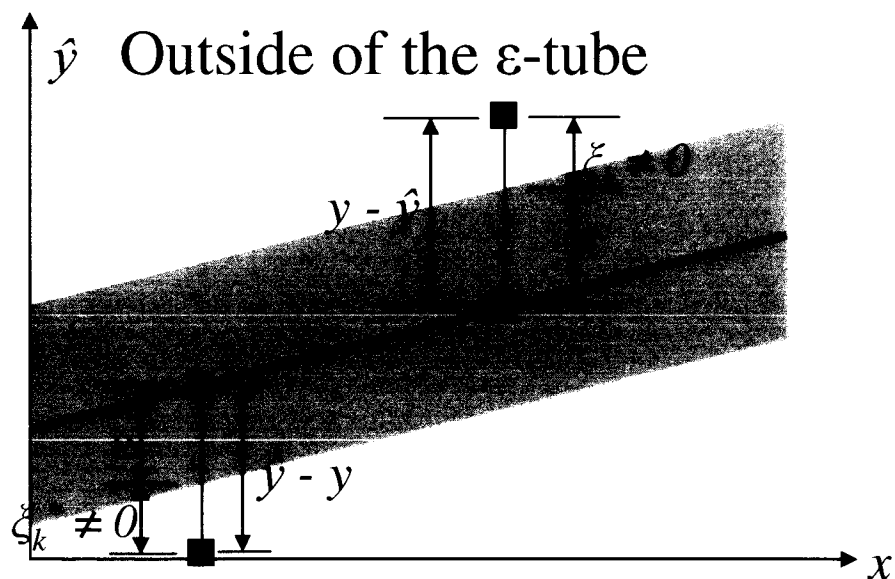
Scalar Case: $y \in \mathbb{R}$

$$\alpha_i = C$$

$$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) - b \triangleq e_i > \varepsilon$$



$$\Rightarrow y_i - \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) - b \triangleq e_i < -\varepsilon$$

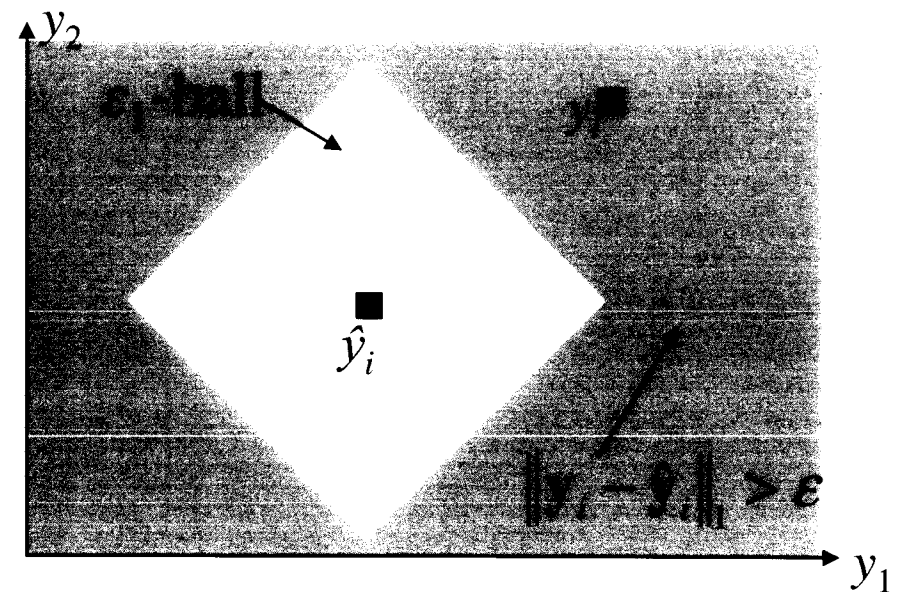


Vector Case: $\mathbf{y} \in \mathbb{R}^m$

$$\alpha_i = \|\boldsymbol{\Gamma}_i\|_q = C$$

$$\Rightarrow \|\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^*\|_p = \|\mathbf{e}_i\|_p > \varepsilon$$

Out of the ε_p -ball



Finding the Bias

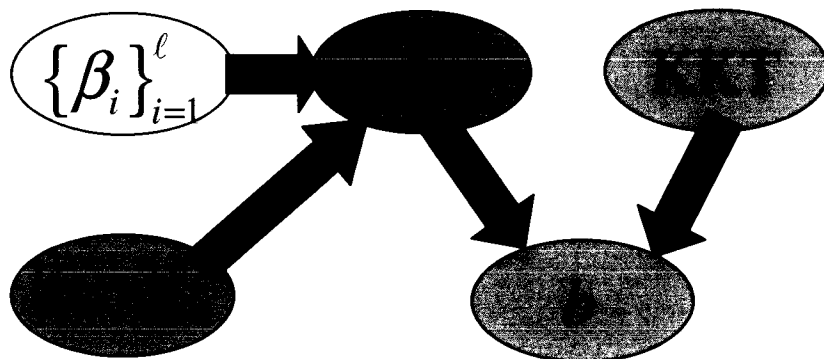
Scalar Case: $y \in \mathbb{R}$

Maximize: $D(\{\beta_i\}_{i=1}^\ell)$ Solve for β_i numerically

Subject to: $\sum_{i=1}^\ell \beta_i = 0, \quad |\beta_i| \leq C$

$$e_i = \mathbf{y}_i - \hat{y}_i$$

$$= \mathbf{y}_i - \underbrace{\sum_{j=1}^\ell \beta_j k(\mathbf{x}_j, \mathbf{x}_i)}_{F_i} - b = \mathbf{y}_i - b$$



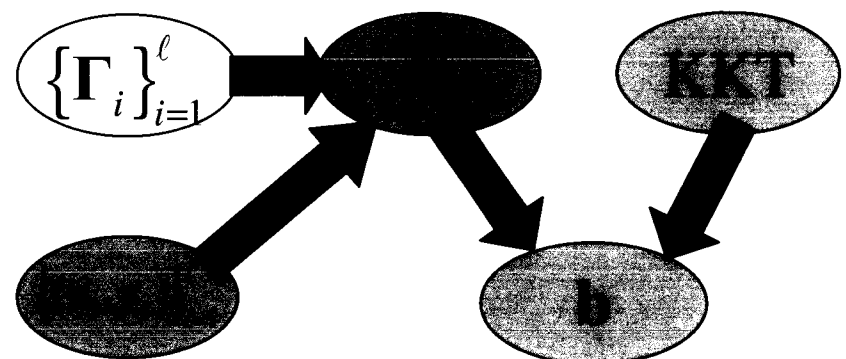
Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Maximize: $D(\{\Gamma_i\}_{i=1}^\ell)$ Solve for Γ_i numerically

Subject to: $\sum_{i=1}^\ell \Gamma_i = 0, \quad \|\Gamma_i\|_q \leq C$

$$\mathbf{e}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i$$

$$= \mathbf{y}_i - \underbrace{\sum_{j=1}^\ell \Gamma_j k(\mathbf{x}_j, \mathbf{x}_i)}_{\mathbf{F}_i} - \mathbf{b} = \mathbf{y}_i - \mathbf{b}$$



Finding the bias ($1 < p < \infty$)

$$\frac{1}{p} + \frac{1}{q} = 1$$

Scalar Case: $y \in \mathbb{R}$

For $i \in \mathcal{M} = \{i : |\beta_i| \in (0, C)\}$

$$F_i - b = e_i = \text{sign}(\beta_i) \varepsilon \quad \text{by KKT}$$

$$\text{if } s_i = \text{sign}(\beta_i)$$

it follows that

$$\boxed{b = F_i - s_i \varepsilon}$$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

For $i \in \mathcal{M} \triangleq \{i : \|\Gamma_i\|_q \in (0, C)\}$

$$\mathbf{F}_i - \mathbf{b} = \mathbf{e}_i = \left(\frac{|\Gamma_i|}{\|\Gamma_i\|_q} \right)^{q-1} \text{sign}(\Gamma_i) \varepsilon \quad \text{by KKT}$$

$$\text{if } \sigma_i = \text{sign}(\Gamma_i)$$

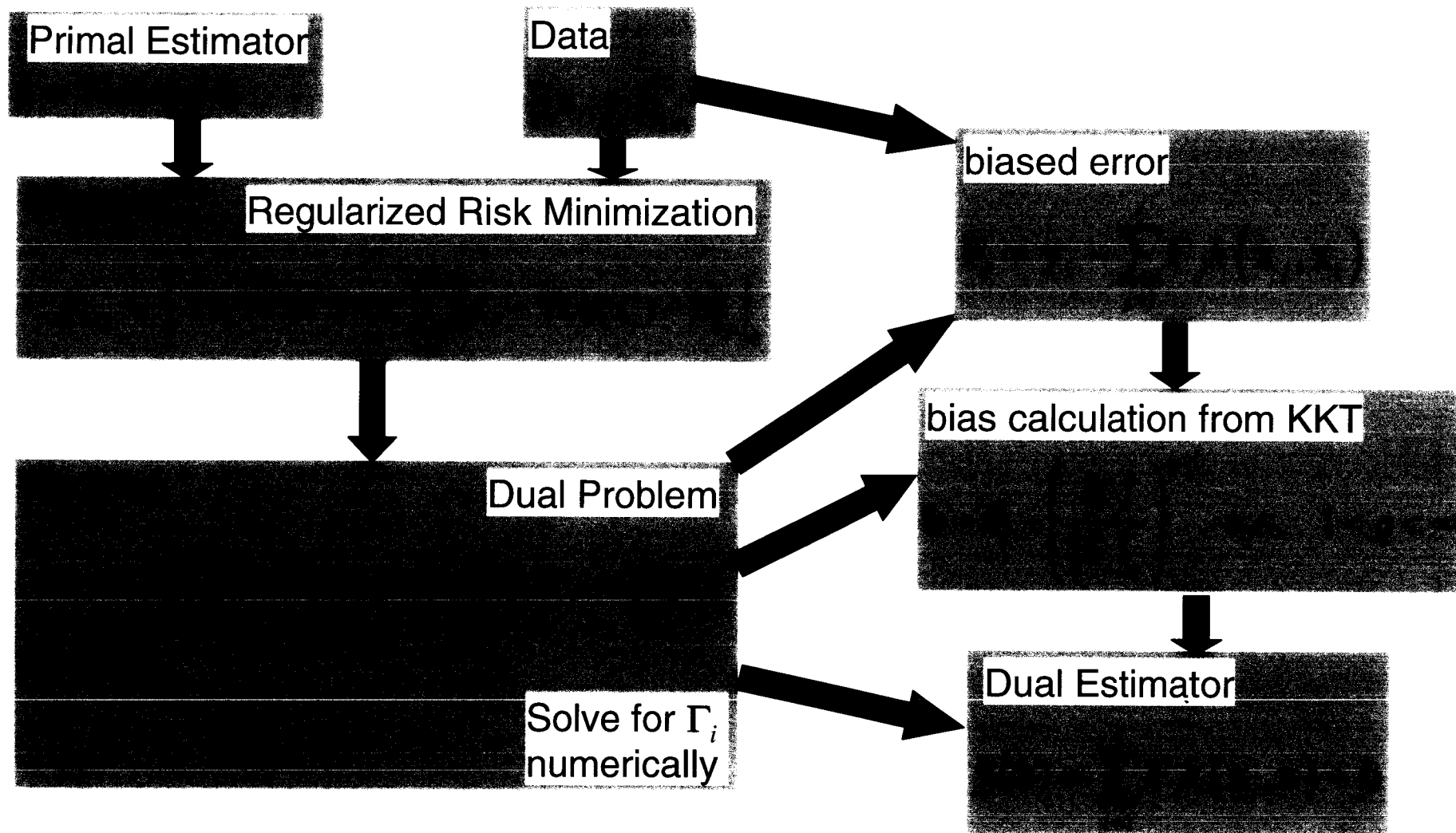
it follows that

$$\boxed{\mathbf{b} = \mathbf{F}_i - \left(\frac{|\Gamma_i|}{\|\Gamma_i\|_q} \right)^{q-1} \sigma_i \varepsilon, \quad 1 < q < \infty}$$

Summary

$$\frac{1}{p} + \frac{1}{q} = 1$$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$



Comparison

Scalar Case: $y \in \mathbb{R}$

Estimator form:

$$\hat{y}(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) + b = \sum_{i \in \mathcal{SV}} \beta_i k(\mathbf{x}_i, \mathbf{x}) + b$$

Optimization Problem:

Maximize:

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \beta_i \beta_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} y_i \beta_i - \varepsilon \sum_{i=1}^{\ell} |\beta_i|$$

$$\text{Subject to: } \sum_{i=1}^{\ell} \beta_i = 0, \quad |\beta_i| \leq C$$

KKT:

$$\begin{aligned} \beta_i = 0 & \rightarrow |e_i| < \varepsilon \\ |\beta_i| \in (0, C) & \rightarrow |e_i| = \varepsilon \\ |\beta_i| = C & \rightarrow |e_i| > \varepsilon \end{aligned}$$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

Estimator form:

$$\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{W} \boldsymbol{\varphi}(\mathbf{x}) + \mathbf{b} = \sum_{i \in \mathcal{SV}} \boldsymbol{\Gamma}_i k(\mathbf{x}_i, \mathbf{x}) + \mathbf{b}$$

Optimization Problem:

Maximize:

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \boldsymbol{\Gamma}_i^T \boldsymbol{\Gamma}_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} \mathbf{y}_i^T \boldsymbol{\Gamma}_i - \varepsilon \sum_{i=1}^{\ell} \|\boldsymbol{\Gamma}_i\|_q$$

$$\text{Subject to: } \sum_{i=1}^{\ell} \boldsymbol{\Gamma}_i = 0, \quad \|\boldsymbol{\Gamma}_i\|_q \leq C$$

KKT:

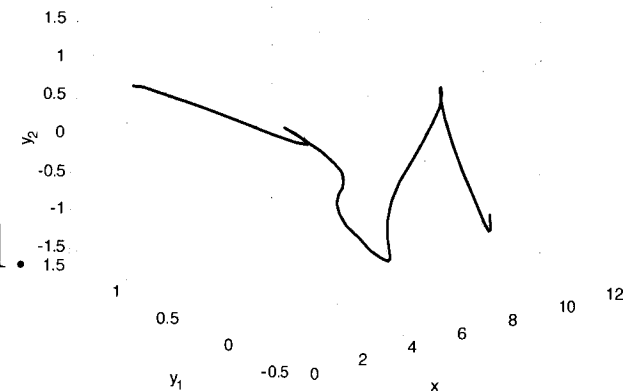
$$\begin{aligned} \|\boldsymbol{\Gamma}_i\|_q = 0 & \rightarrow \|\mathbf{e}_i\|_p < \varepsilon \\ \|\boldsymbol{\Gamma}_i\|_q \in (0, C) & \rightarrow \|\mathbf{e}_i\|_p = \varepsilon \\ \|\boldsymbol{\Gamma}_i\|_q = C & \rightarrow \|\mathbf{e}_i\|_p > \varepsilon \end{aligned}$$

Demonstrations for $p = 1, 2, \infty$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

We will demonstrate the VV-SVR for the following process:

$$\mathbf{y}(x) = \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} e^{0.1x} \text{sinc}(x) \\ \cos(0.1x^2) \end{bmatrix}$$



The following RBF kernel was used with $\sigma = 1$.

$$k(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{2\sigma^2}\right)$$

50 samples were chosen on the interval $[0, 10]$ for training and $\varepsilon = 0.1$.

Matlab was used as the solver.

The 1-Norm

$p = 1$

$q \sim \infty$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

The dual problem:

Maximize:

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \Gamma_i^T \Gamma_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} \mathbf{y}_i^T \Gamma_i - \varepsilon \sum_{i=1}^{\ell} \|\Gamma_i\|_{\infty}$$

Subject to:

$$\sum_{i=1}^{\ell} \Gamma_i = 0, \quad \|\Gamma_i\|_{\infty} \leq C$$

Note: D is non-smooth in its objective and constraint.

The dual problem after introduction of α_i :

Maximize:

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \Gamma_i^T \Gamma_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} \mathbf{y}_i^T \Gamma_i - \varepsilon \sum_{i=1}^{\ell} \alpha_i$$

Subject to:

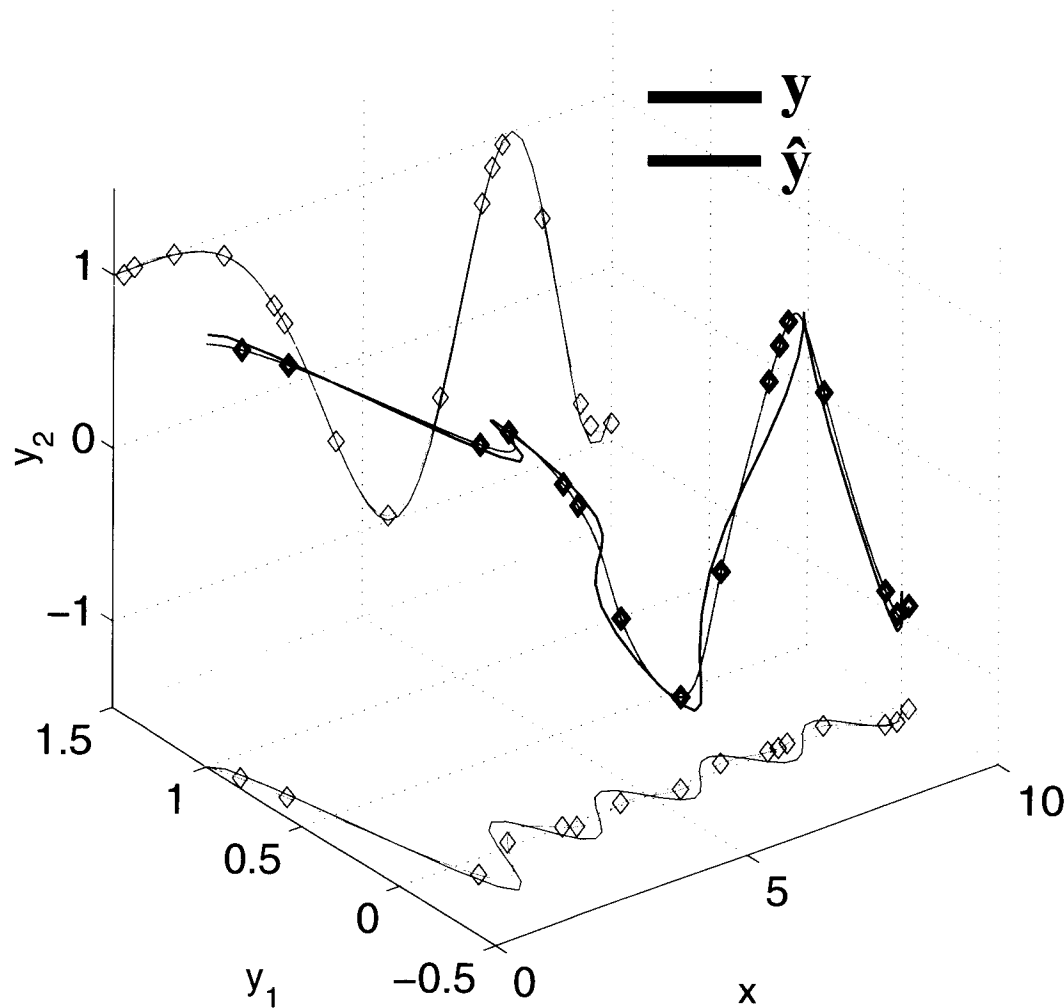
$$\sum_{i=1}^{\ell} \Gamma_i = 0, \quad -\alpha_i \mathbf{1} \leq \Gamma_i \leq \alpha_i \mathbf{1}, \quad \alpha_i \leq C$$

Note: D is quadratic in its objective and linear in its constraint.

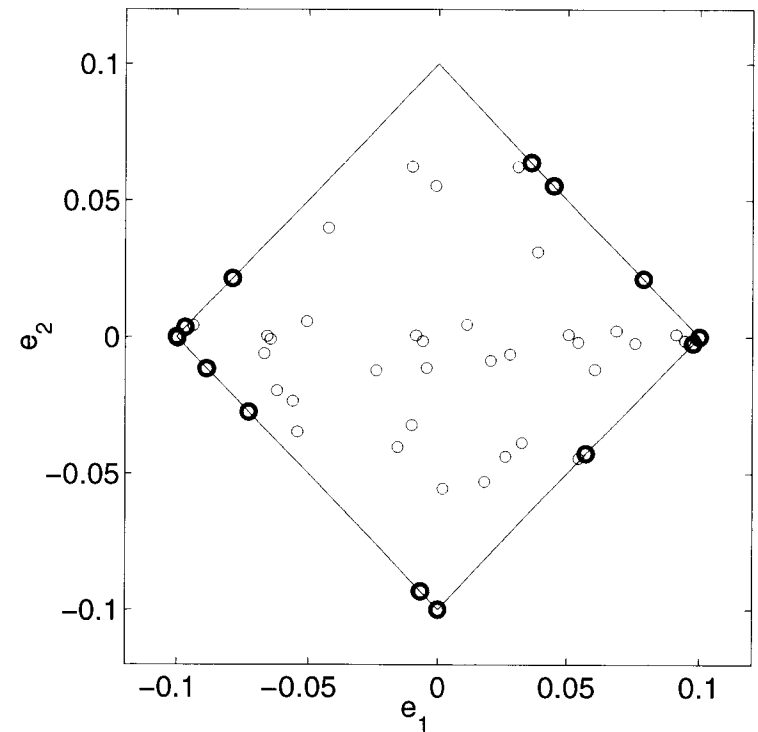
Solve for Γ_i using a standard QP package.

Demonstration (1-Norm)

Vector Case: $y \in \mathbb{R}^m$



16 support vectors



Matlab's `quadprog()`
was used to find Γ_i .

The 2-Norm

$p = 2$
 $q = 2$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

The dual problem:

Maximize:

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \Gamma_i^T \Gamma_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} \mathbf{y}_i^T \Gamma_i - \varepsilon \sum_{i=1}^{\ell} \|\Gamma_i\|_2$$

Subject to:

$$\sum_{i=1}^{\ell} \Gamma_i = 0, \quad \|\Gamma_i\|_2 \leq C$$

Note: D is non-smooth in its objective and constraint.

The constraint is nonlinear.

The dual problem after introduction of α_i :

Maximize:

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \Gamma_i^T \Gamma_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} \mathbf{y}_i^T \Gamma_i - \varepsilon \sum_{i=1}^{\ell} \alpha_i$$

Subject to:

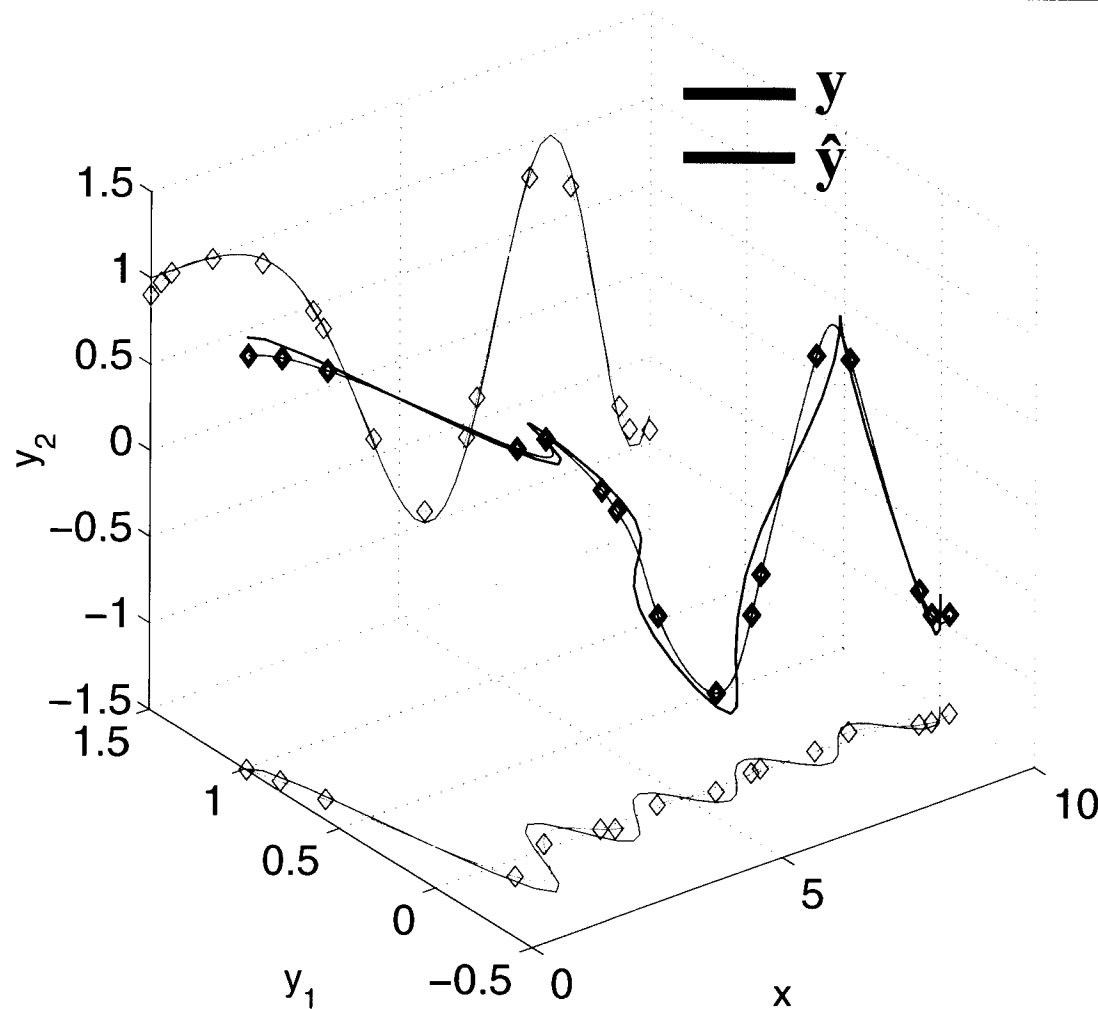
$$\sum_{i=1}^{\ell} \Gamma_i = 0, \quad \Gamma_i^T \Gamma_i = \alpha_i^2, \quad 0 \leq \alpha_i \leq C$$

Note: D is quadratic in its objective and nonlinear but smooth in its constraint.

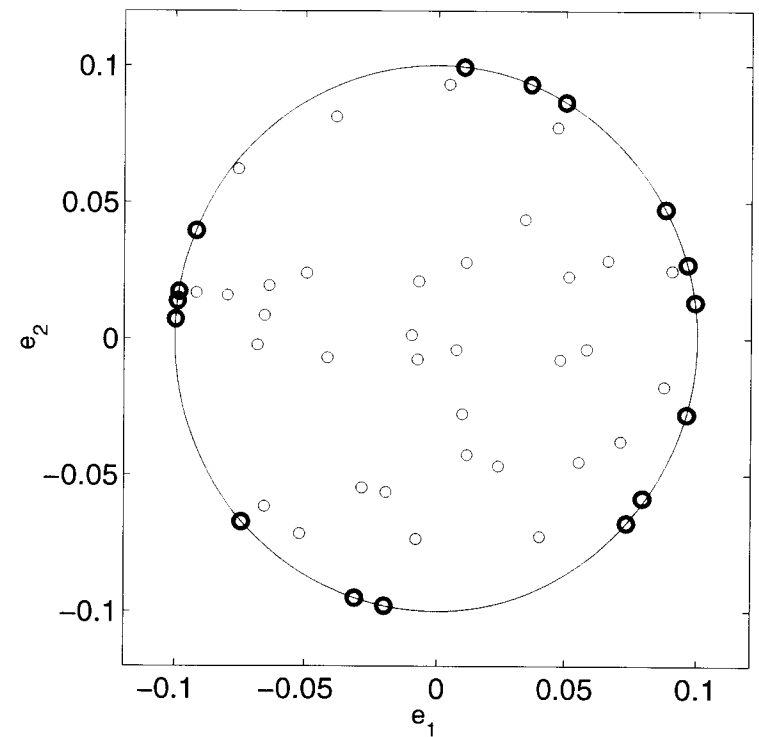
Solve for Γ_i using a standard nonlinear programming package which can use gradients.

Demonstration (2-Norm)

Vector Case: $y \in \mathbb{R}^m$



16 support vectors



Matlab's `fmincon()`
was used to find Γ_i .

The ∞ -Norm

$$\begin{aligned} p &\sim \infty \\ q &= 1 \end{aligned}$$

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

The dual problem:

Maximize:

$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \Gamma_i^T \Gamma_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell} \mathbf{y}_i^T \Gamma_i - \varepsilon \sum_{i=1}^{\ell} \|\Gamma_i\|_1$$

Subject to:

$$\sum_{i=1}^{\ell} \Gamma_i = 0, \quad \|\Gamma_i\|_1 \leq C$$

Note: D is non-smooth in its objective and constraint.

The dual problem after introduction of γ_i and γ_i^* :

Maximize:

$$\begin{aligned} D = & -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (\gamma_i - \gamma_i^*)^T (\gamma_j - \gamma_j^*) k(\mathbf{x}_i, \mathbf{x}_j) \\ & + \sum_{i=1}^{\ell} \mathbf{y}_i^T (\gamma_i - \gamma_i^*) - \varepsilon \sum_{i=1}^{\ell} \mathbf{1}^T (\gamma_i + \gamma_i^*) \end{aligned}$$

Subject to:

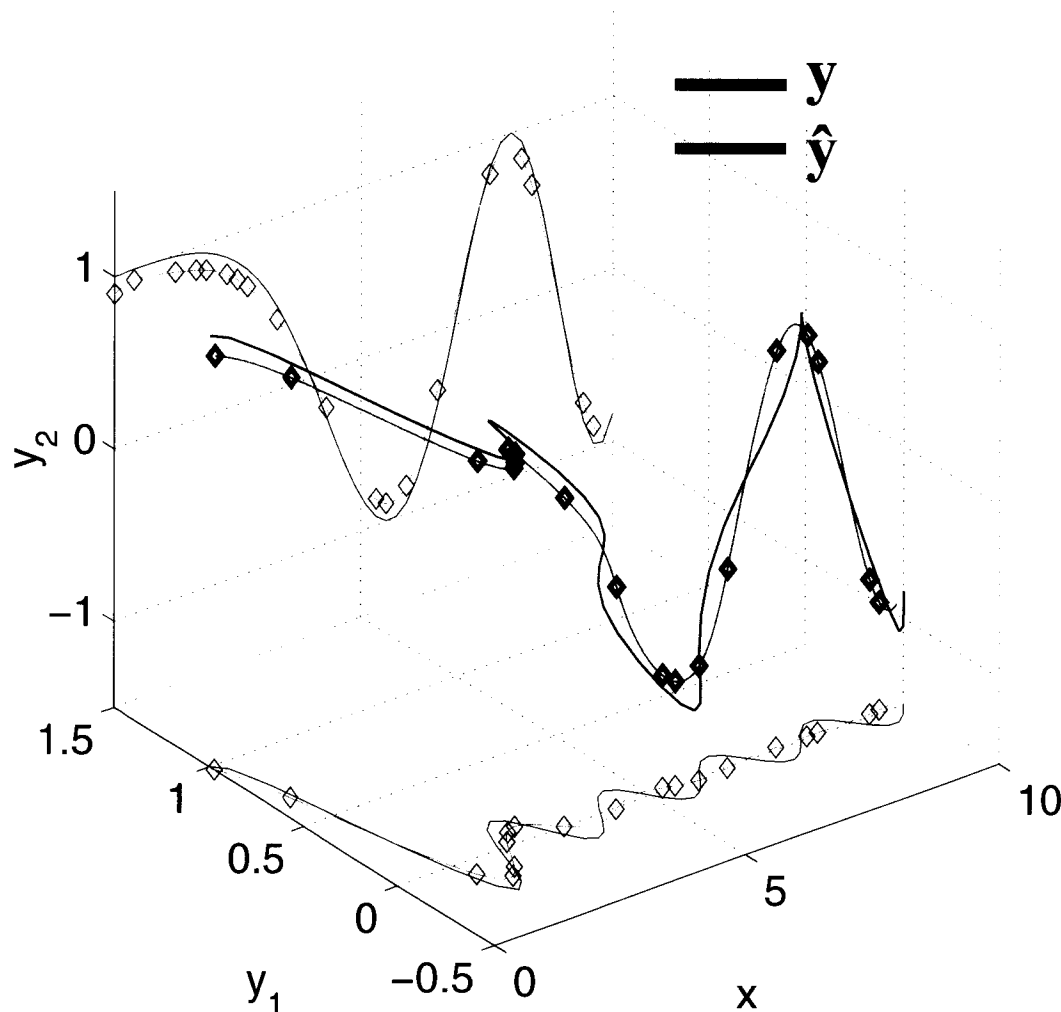
$$\sum_{i=1}^{\ell} \Gamma_i = 0, \quad \mathbf{1}^T (\gamma_i + \gamma_i^*) \leq C, \quad \gamma_i, \gamma_i^* \geq 0$$

Note: D is quadratic in its objective and linear in its constraint.

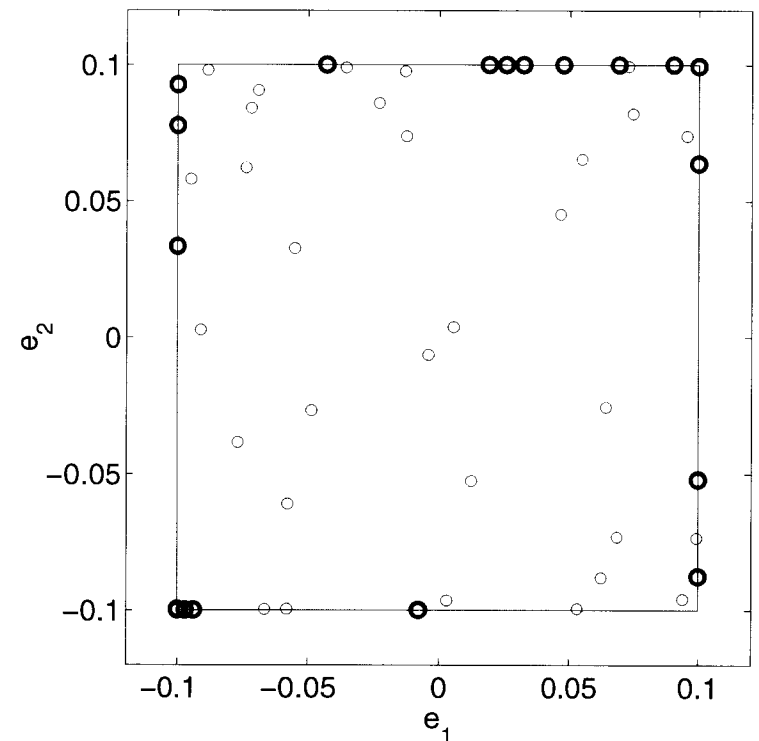
Solve for Γ_i using a standard QP package.

Demonstration (∞ -Norm)

Vector Case: $\mathbf{y} \in \mathbb{R}^m$



19 support vectors



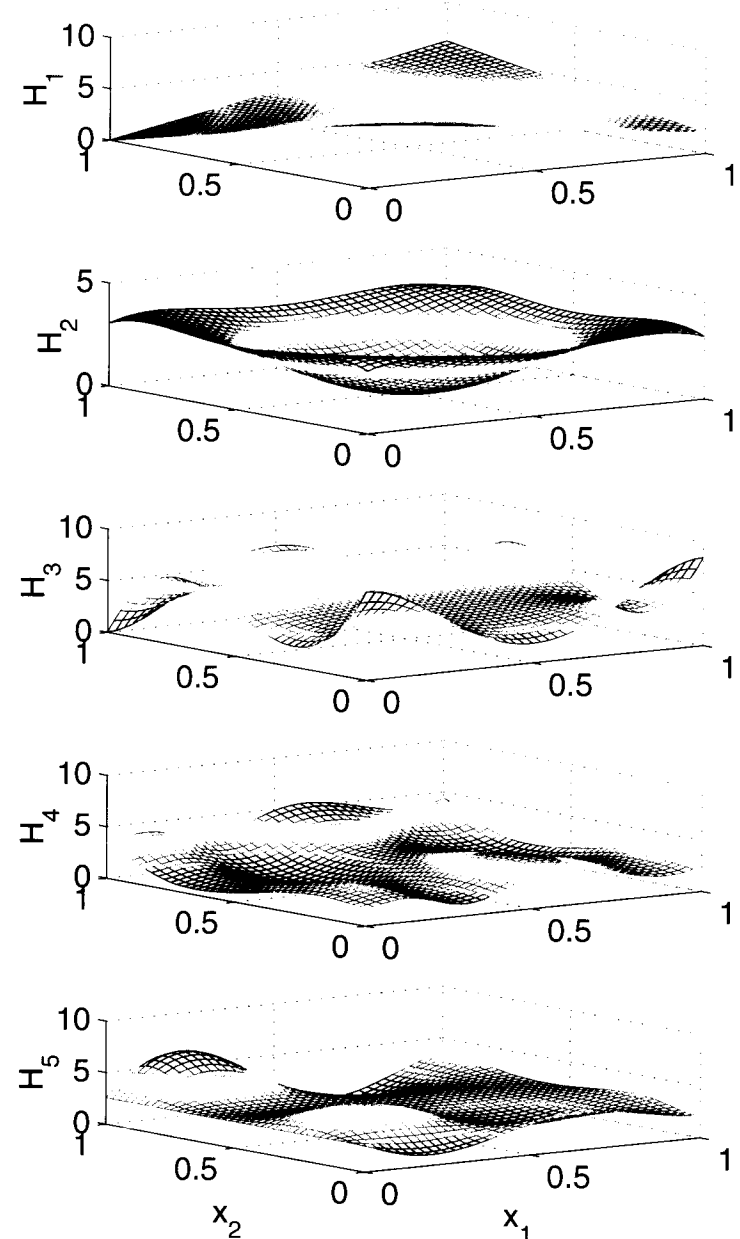
Matlab's `quadprog()` was used to find Γ_i .

Example Hwang

- Hwang data set

$$\mathbf{H} : [0,1]^2 \mapsto \mathbb{R}_+^5$$

- Artificial vector-valued data set
- Of Historical significance
- Input domain is randomly sampled



Hwang

$$\ell = 500, \quad \varepsilon = 0.5$$

$$C = 100, \quad \gamma = 8$$

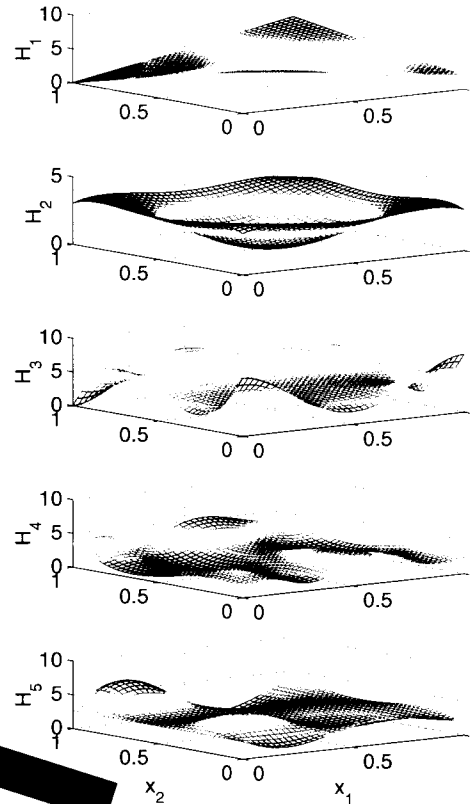
$$\{\mathbf{x}_i, \mathbf{H}_i\}_{i=1}^{\ell}$$

Maximize:

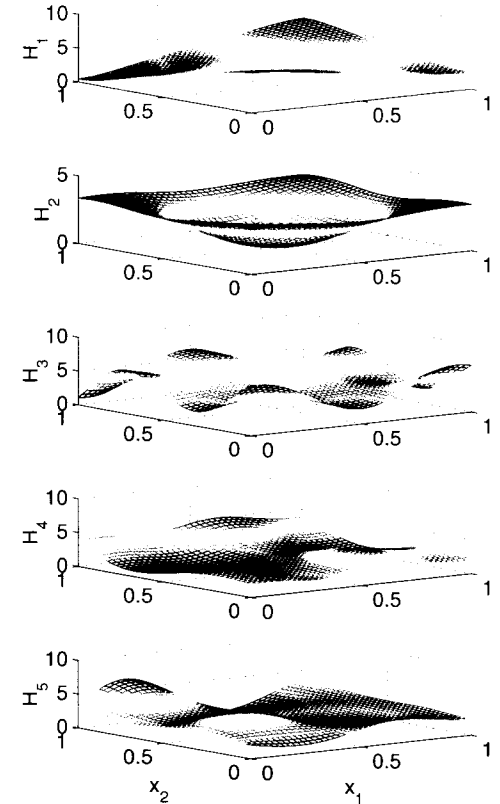
$$D = -\frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \Gamma_i^T \Gamma_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$+ \sum_{i=1}^{\ell} \mathbf{H}_i^T \Gamma_i - \varepsilon \sum_{i=1}^{\ell} \|\Gamma_i\|_2$$

Subject to: $\sum_{i=1}^{\ell} \Gamma_i = 0, \quad \|\Gamma_i\|_2 \leq C$



$\mathbf{H}(\mathbf{x})$

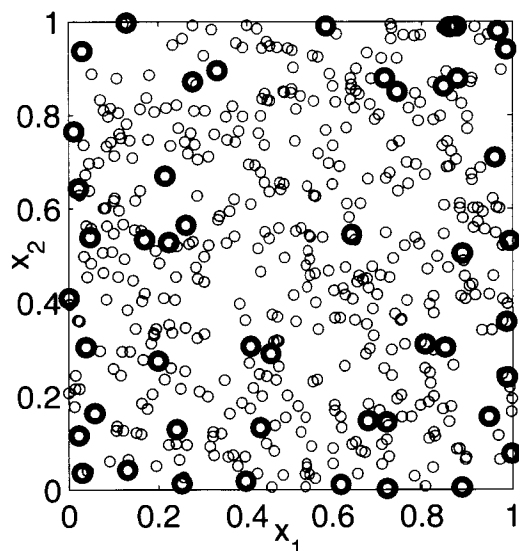


$\hat{\mathbf{H}}(\mathbf{x})$

$$\hat{\mathbf{H}}(\mathbf{x}) = \sum_{i \in \mathcal{SV}} \Gamma_i k(\mathbf{x}_i, \mathbf{x}) + \mathbf{b}$$

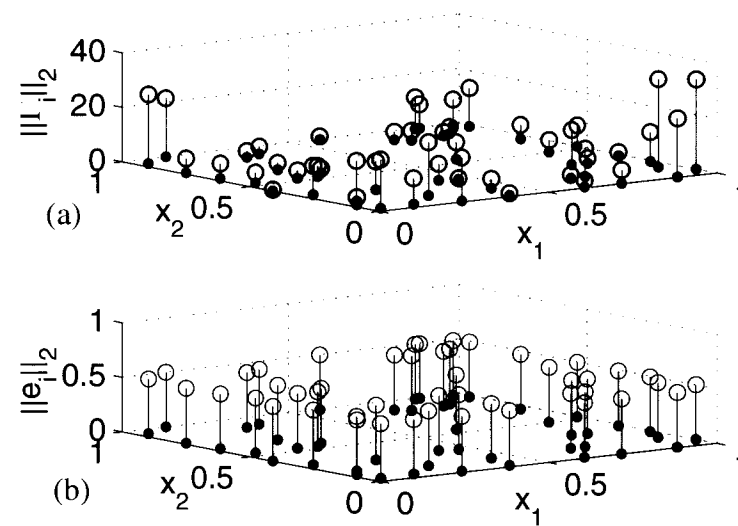
$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2\right)$$

Hwang KKT



48 Support Vectors

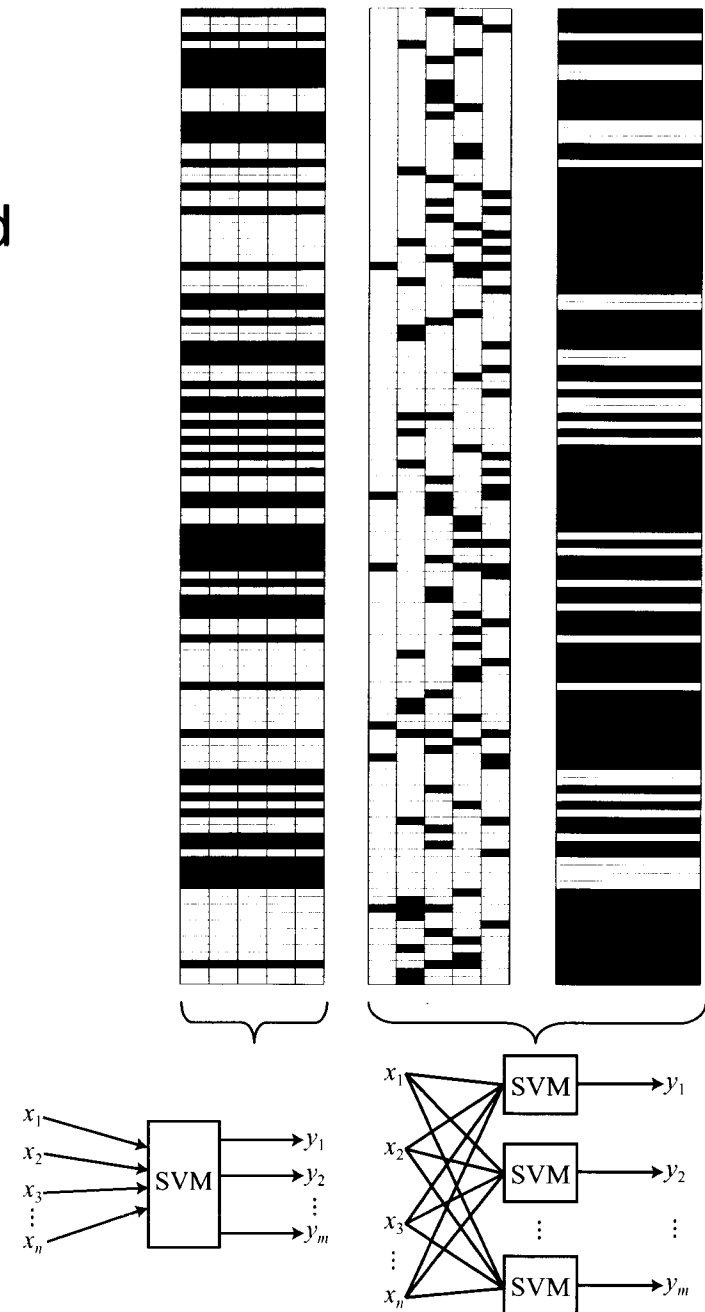
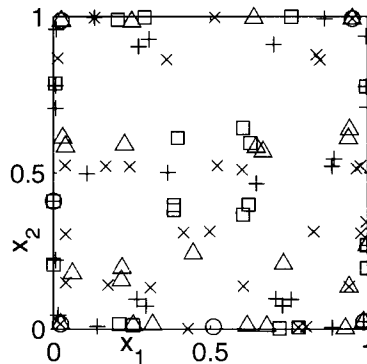
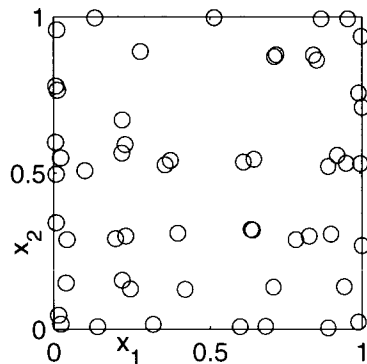
$$\|\Gamma_i\|_2 \in (0, C) \Rightarrow \|\mathbf{e}_i\|_2 = \varepsilon$$



KKT Conditions

Sparsisity

- Compare VV-SVR with aggregated SVR (libsvm)
- Hwang data set with 2,000 points
- Equal Volume
 - VV-SVR $\varepsilon = 0.5$
 - SVR $\varepsilon = 0.34850$
- Support Vectors
 - VV-SVR 55
 - SVR 92



Conclusions

Vector Case: $\mathbf{y} \in \mathbb{R}^m$

- VV-SVR generalizes the scalar-valued case.
 - Estimator form and parameters
 - Loss function
 - Regularization functional
- VV-SVR maintains the sparsity of the scalar-valued case.