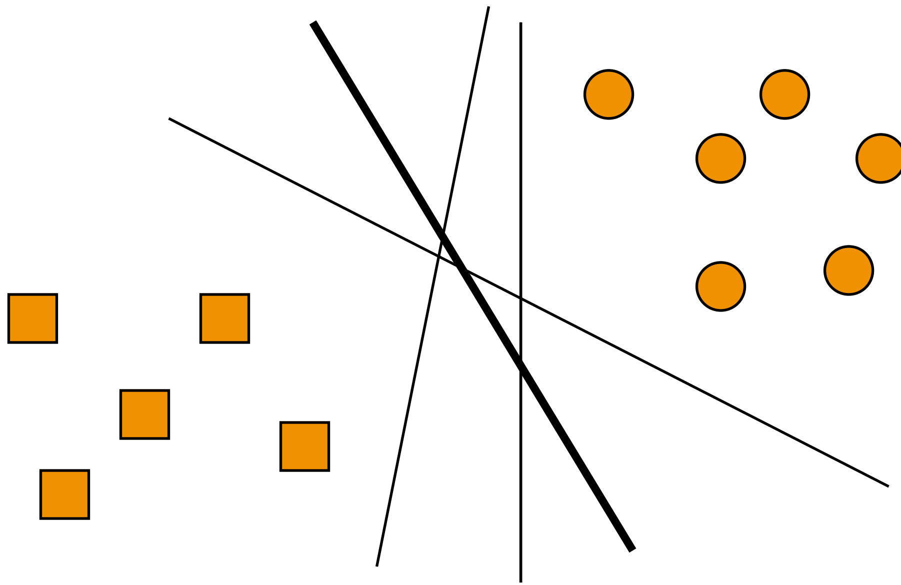
Support Vector Machine Classification Technique

Quang-Anh Tran
Email: qa00@mails.tsinghua.edu.cn
Department of Electronic Engineering
Tsinghua University, China

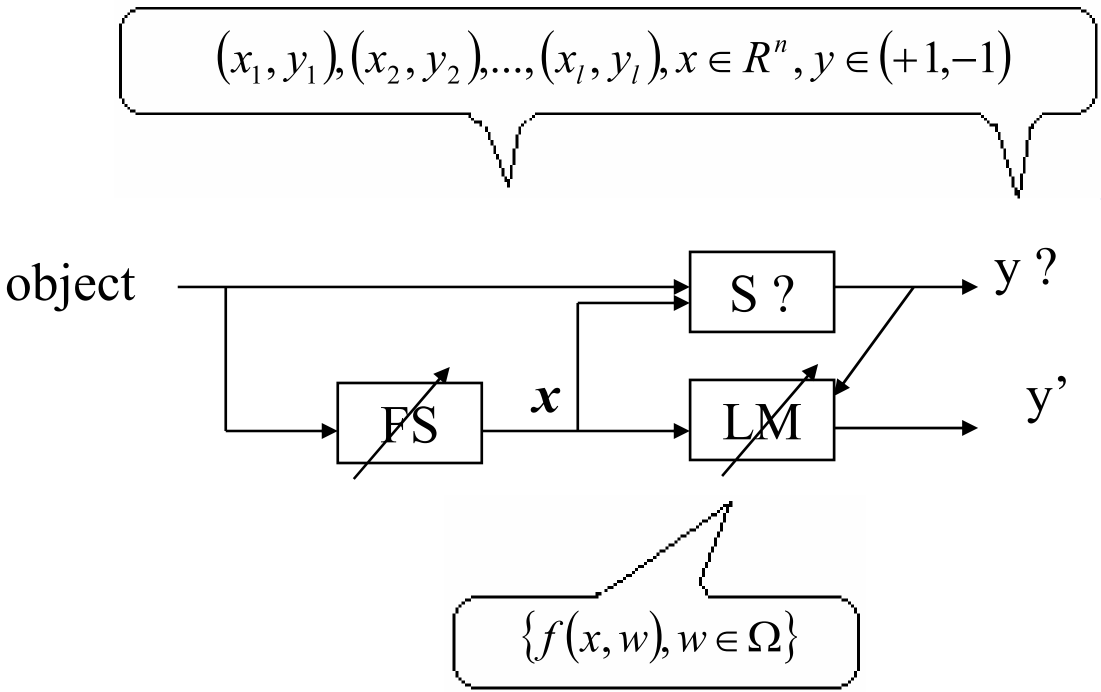
2003-2-28
Presentation at the Lab of New Generation Network, Tsinghua University

Problem



Contents

1. Statistical learning theory



1. Statistical learning theory

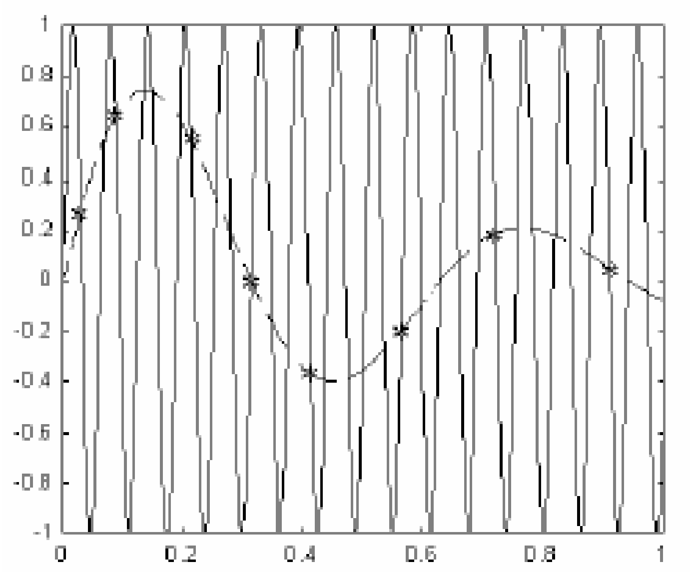
- Risk

$$R(w) = \int L\big(y, f(\mathbf{x}, w)\big) dP(\mathbf{x}, y)$$

- Empirical risk

$$R_{emp}(w) = \frac{1}{l}\sum_{i=1}^{l} L\big(y_i, f(\mathbf{x}_i, w)\big)$$

1. Statistical learning theory



1. Statistical learning theory

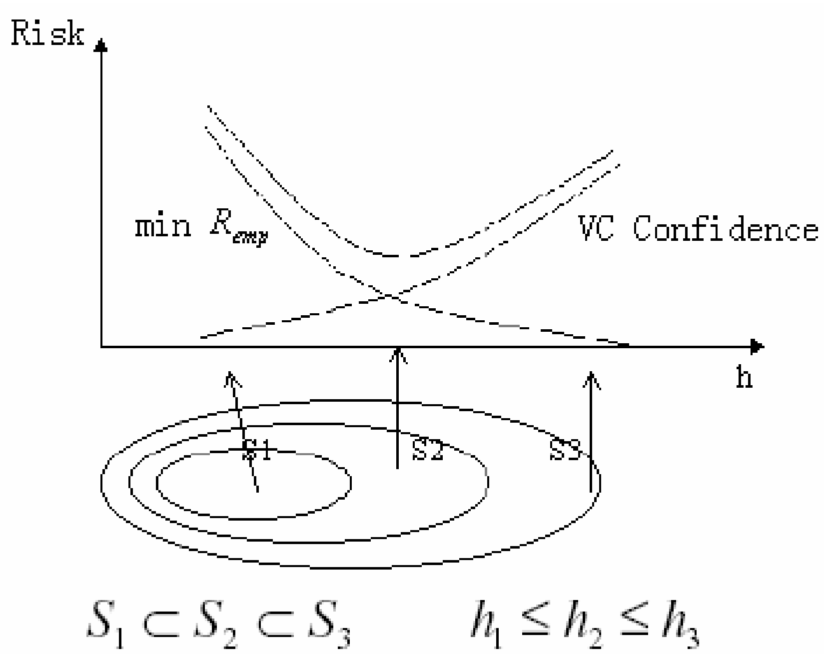
- The following inequality is valid with probability $1-\eta$

$$R(w) \le R_{emp}(w) + \sqrt{\frac{h(\ln(2l/h)+1)-\ln(\eta/4)}{l}}$$

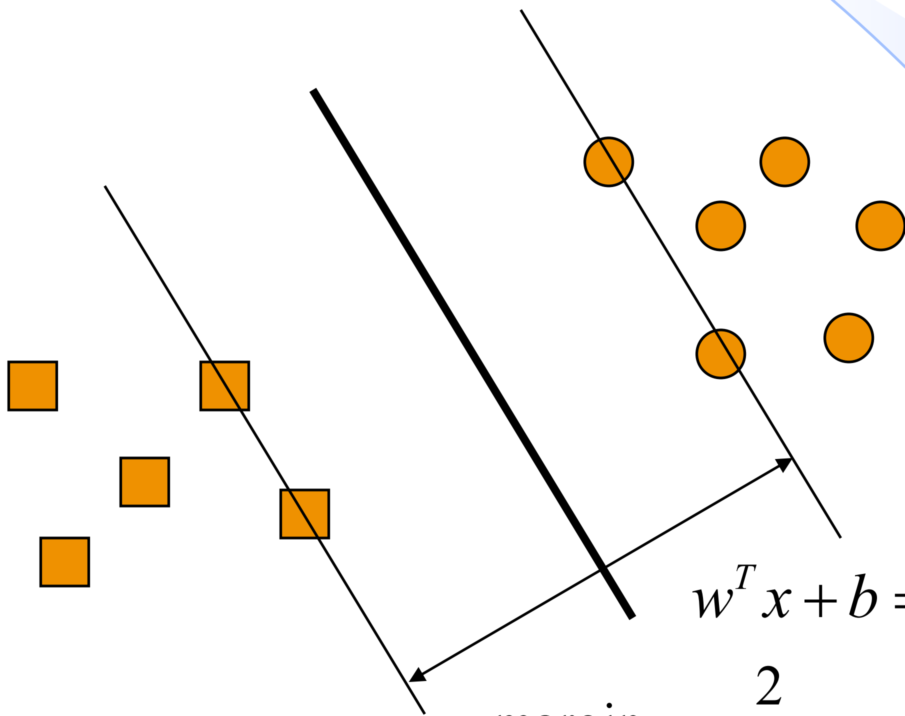
VC Confidence

1. Statistical learning theory

- Structure Risk Minimization – SRM



2. Maximal margin hyperplane



$$w^T x + b = 0$$

$$\text{margin} = \frac{2}{\|w\|^2}$$

2. Maximal margin hyperplane

- The SRM in SVM:

$$w^T x + b = 0 \qquad \|w\| = 1$$

$$S_\Delta : \qquad y = \begin{cases} +1, & if\left(w^T x + b \ge \Delta\right) \\ -1, & if\left(w^T x + b \le -\Delta\right) \end{cases}$$

$$h_\Delta \le \min\left(\left[\frac{R^2}{\Delta^2}\right], n\right) + 1$$

2. Maximal margin hyperplane

- The data is linearly separable

$$\min_{w} \frac{1}{2} w^T w$$

Subject to:

$$y_i\left(w^T x_i + b\right) \geq 1, i = 1, \ldots, l$$

$$\Rightarrow$$

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

Subject to:

$$0 \leq \alpha_i, i = 1, \ldots, l$$

$$y^T \alpha = 0$$

$$Q_{ij} = y_i y_j x_i^T x_j$$

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i x_i^T x + b\right)$$

2. Maximal margin hyperplane

- The data is linearly separable

$$\min_{w,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

Subject to:

$$y_i \left(w^T x_i + b \right) \geq 1 - \xi_i, i = 1, \ldots, l$$

$$\Rightarrow$$

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

Subject to:

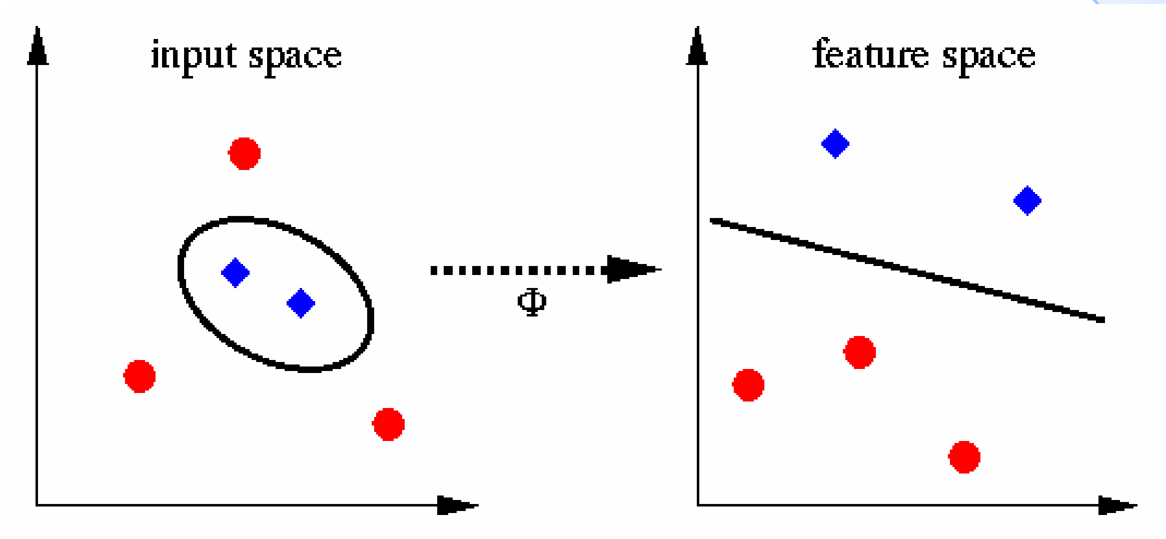
$$0 \leq \alpha_i \leq C, i = 1, \ldots, l$$

$$y^T \alpha = 0$$

$$Q_{ij} = y_i y_j x_i^T x_j$$

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i x_i^T x + b \right)$$

3. Kernel methods



3. Kernel methods

$$\min_{w,\xi} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

Subject to:

$$y_i \left(w^T \Phi(x_i) + b \right) \geq 1 - \xi_i, i = 1, \ldots, l$$

$$\Rightarrow$$

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

Subject to:

$$0 \leq \alpha_i \leq C, i = 1, \ldots, l$$

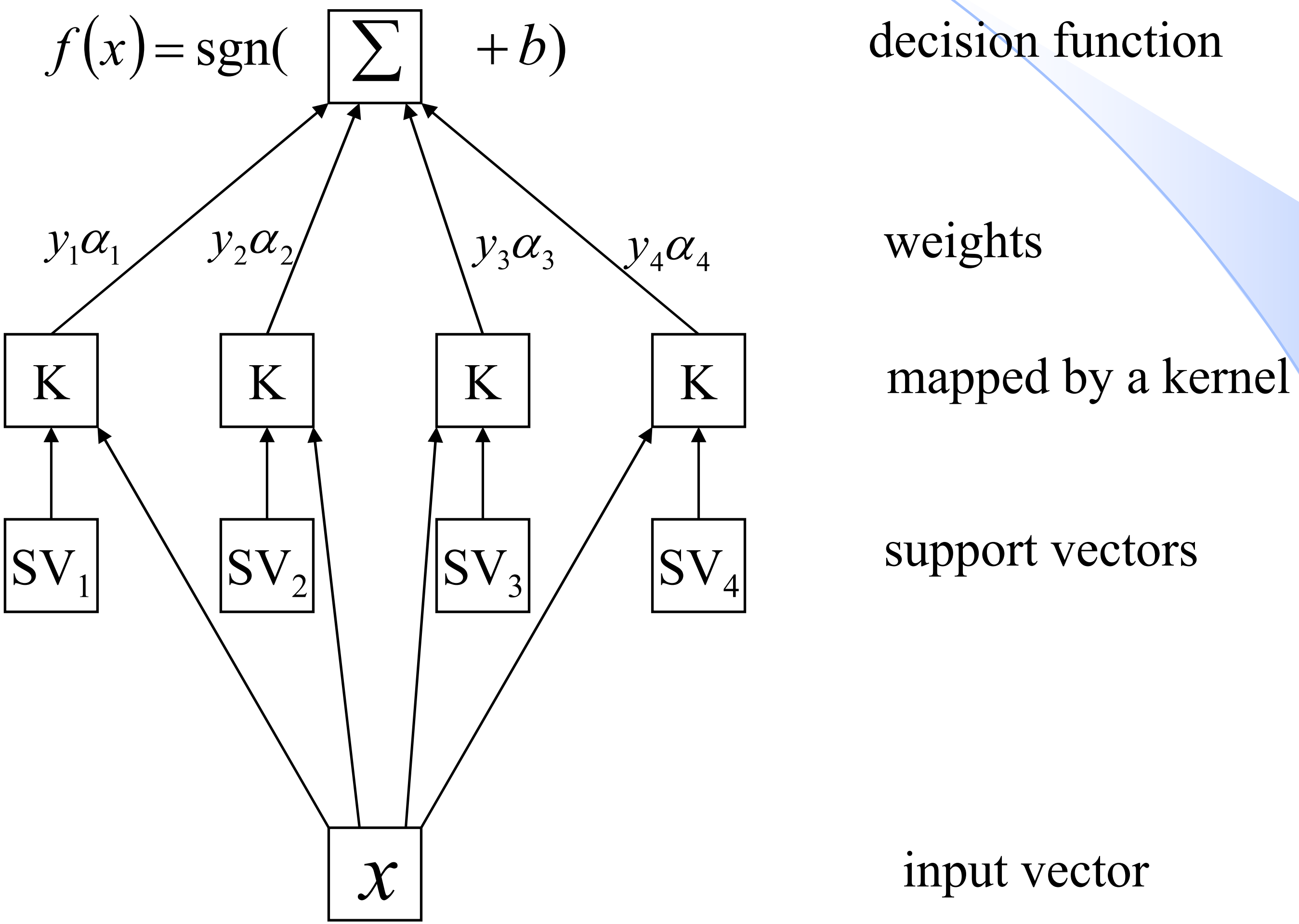
$$y^T \alpha = 0$$

$$Q_{ij} = y_i y_j \Phi(x_i)^T \Phi(x_j)$$

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i \Phi(x_i)^T \Phi(x) + b \right)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$$

3. Kernel methods



3. Kernel methods

- Polynomial

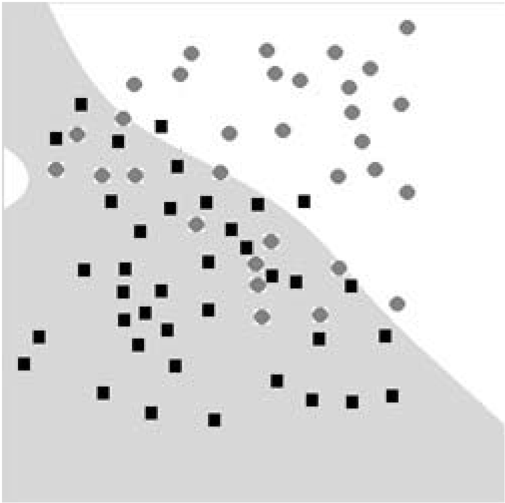
$$K_{poly}(x_1, x_2) = \left(s{x_1}^T x_2 + r\right)^d$$

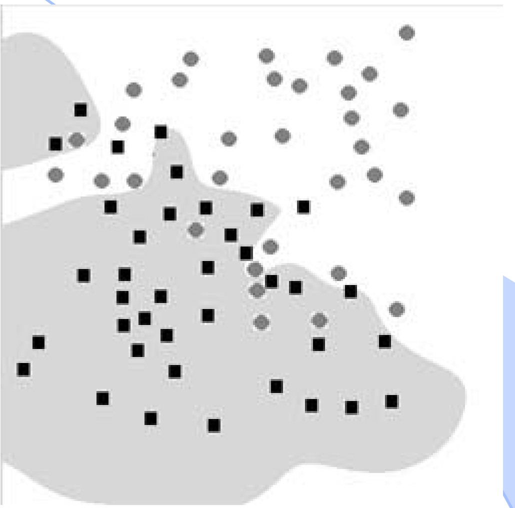
- Radial Basic Function (RBF)

$$K_{RBF}\left(x_1, x_2\right) = e^{\frac{\left\|x_1 - x_2\right\|^2}{2\sigma^2}}$$

3. Kernel methods

K_{poly}





K_{RBF}

$$K_{mix} = \lambda K_{RBF} + (1-\lambda) K_{poly}$$

$$0 \leq \lambda \leq 1$$

4. Types of SVM

- C-SVC: C_+, C-
- ν-SVC: ν , $C = 1/\rho$
- One-class SVM: ν
- Transductive SVM: semi-supervised
- Invariant SVM: includes pre-processing

4. Types of SVM

- C-SVC

$$\min_{w,\xi} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$

Subject to:

$$y_i \left(w^T x_i + b \right) \geq 1 - \xi_i, i = 1, \ldots, l$$

$$\Rightarrow$$

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

Subject to:

$$0 \leq \alpha_i \leq C, i = 1, \ldots, l$$

$$y^T \alpha = 0$$

$$Q_{ij} = y_i y_j x_i^T x_j$$

4. Types of SVM

$$\min_{w,\xi} \frac{1}{2} w^T w + C_+ \sum_{i=1}^{l_1} \xi_i + C_- \sum_{i=1}^{l_2} \xi_i$$

Subject to:

$$w^T x_i + b \geq 1 - \xi_i, i = 1, \ldots, l_1$$

$$w^T x_i + b \leq -1 + \xi_i, i = 1, \ldots, l_2$$

$$\Rightarrow$$

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

Subject to:

$$0 \leq \alpha_i \leq C_+, i = 1, \ldots, l_1$$

$$0 \leq \alpha_i \leq C_-, i = 1, \ldots, l_2$$

$$y^T \alpha = 0$$

$$Q_{ij} = y_i y_j x_i^T x_j$$

4. Types of SVM

- ν-SVC:

$$\min_{\omega,\rho,\xi} \quad \frac{1}{2}\omega^T\omega - \nu\rho + \frac{1}{l}\sum_{i=1}^{l}\xi_i$$

Subject to:

$$y_i\left(\omega^T\phi(x_i) + b\right) \geq \rho - \xi_i$$

$$\xi_i \geq 0, i = 1,\ldots,l, \rho \geq 0$$

$$\Rightarrow$$

$$\min_{\alpha} \quad \frac{1}{2}\alpha^T Q\alpha$$

Subject to:

$$0 \leq \alpha_i \leq \frac{1}{l}, i = 1,\ldots,l$$

$$e^T\alpha \geq \nu$$

$$y^T\alpha = 0$$

$$Q_{ij} = y_i y_j x_i^T x_j$$

4. Types of SVM

- One-class SVM

$$\min_{\omega,\xi,\rho} \quad \frac{1}{2}\omega^T\omega - \rho + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i$$

Subject to:

$$\omega^T\phi(x_i) \geq \rho - \xi_i$$

$$\xi_i \geq 0, i = 1,\ldots,l$$

$$\Rightarrow$$

$$\min_{\alpha} \quad \frac{1}{2}\alpha^T Q\alpha$$

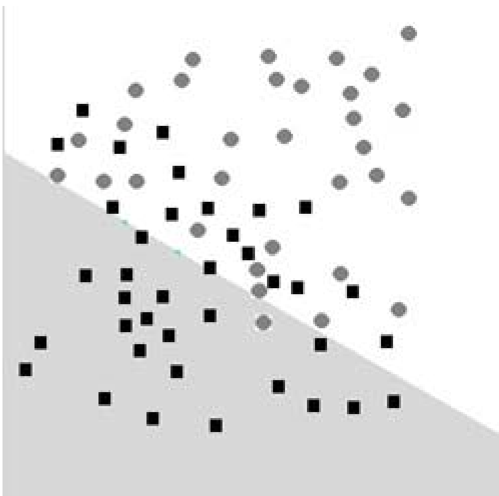
Subject to:

$$0 \leq \alpha_i \leq \frac{1}{\nu l}, i = 1,\ldots,l$$

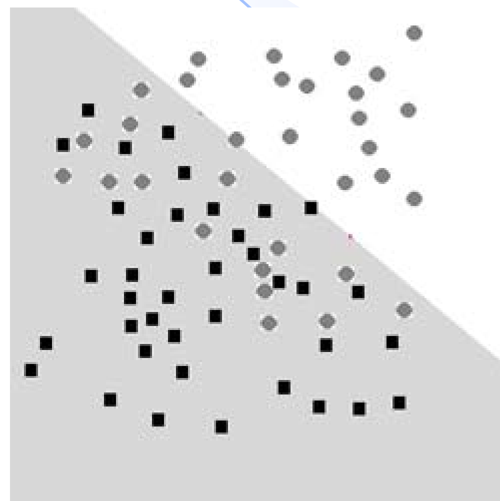
$$e^T\alpha = 1$$

$$Q_{ij} = y_i y_j x_i^T x_j$$

4. Types of SVM



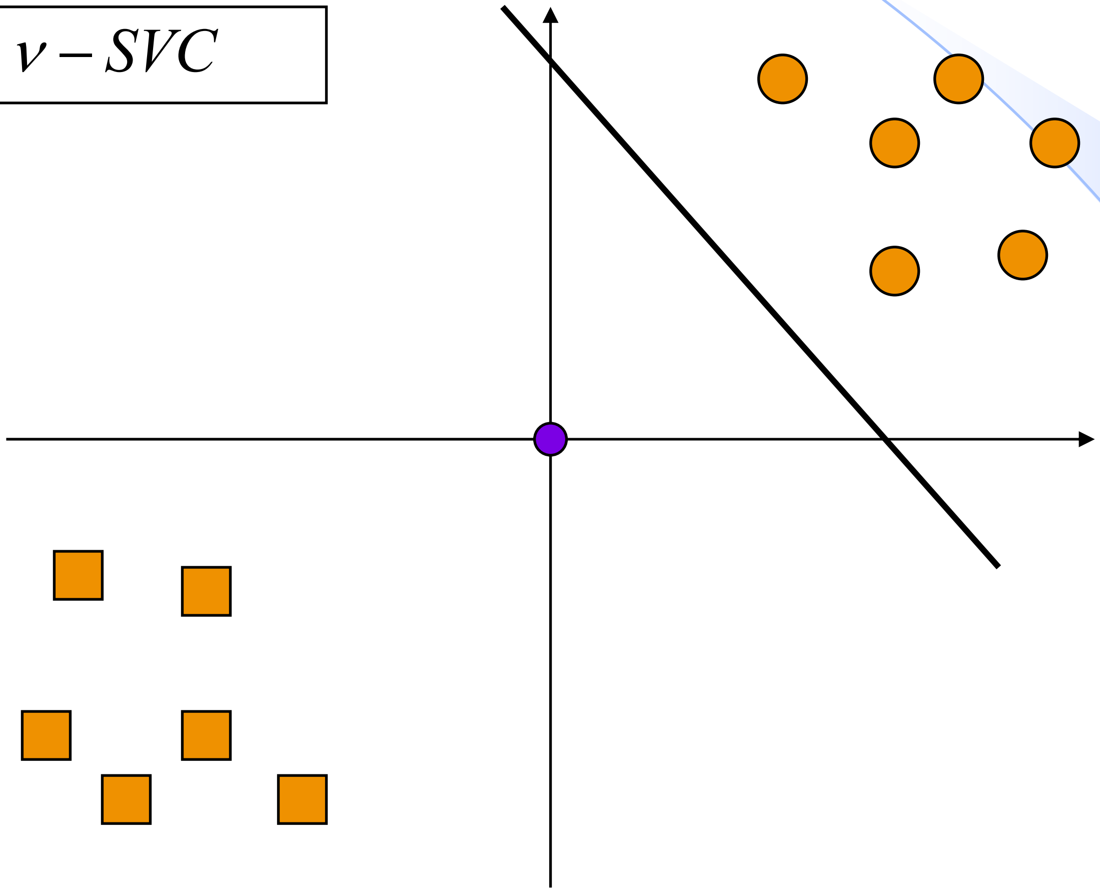
$C_+ = 1, C_- = 10$



$C_+ = 10, C_- = 1$

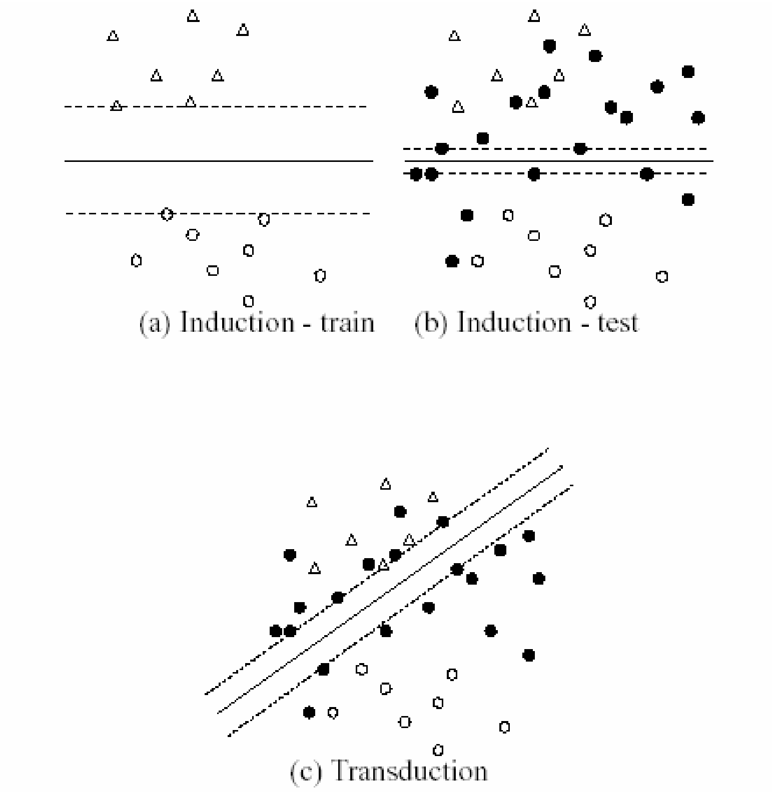
4. Types of SVM

$\nu - SVC$



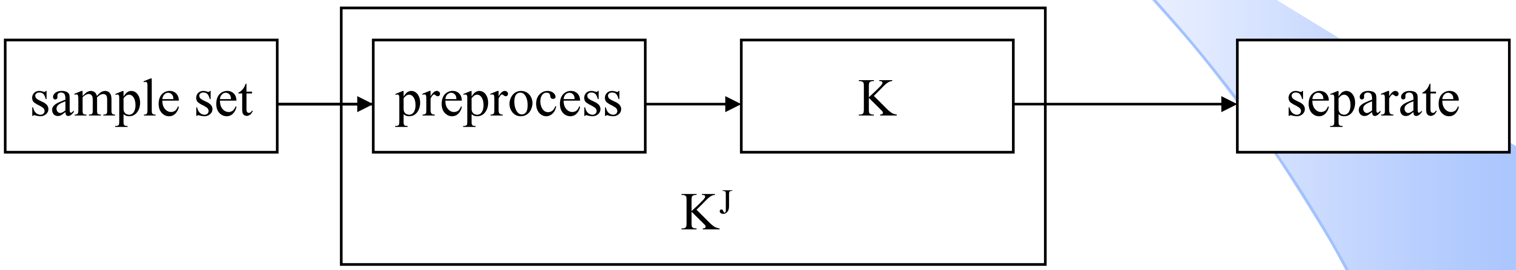
4. Types of SVM

Transductive SVM



4. Types of SVM

Invariant SVM



5. Implementation

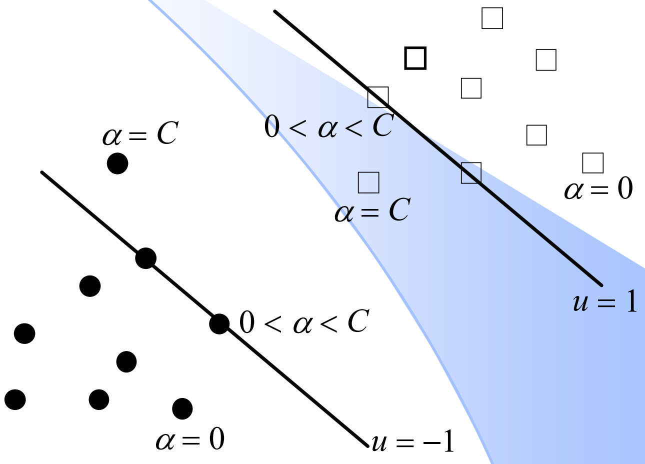
$$\min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - e^T\alpha$$

Subject to:

$$0 \le \alpha_i \le C, i = 1,\ldots,l$$

$$y^T\alpha = 0$$

$$Q_{ij} = y_i y_j x_i^T x_j$$



Kühn-Tucker Condition

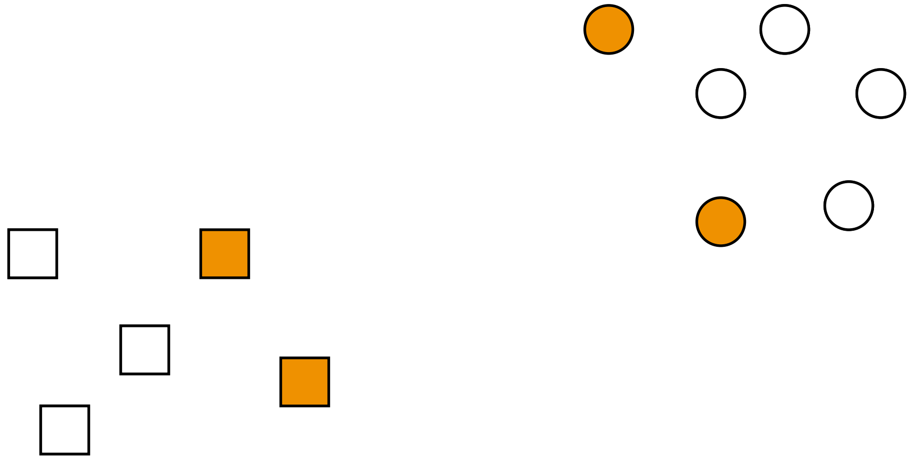
Decomposition method

5. Implementation

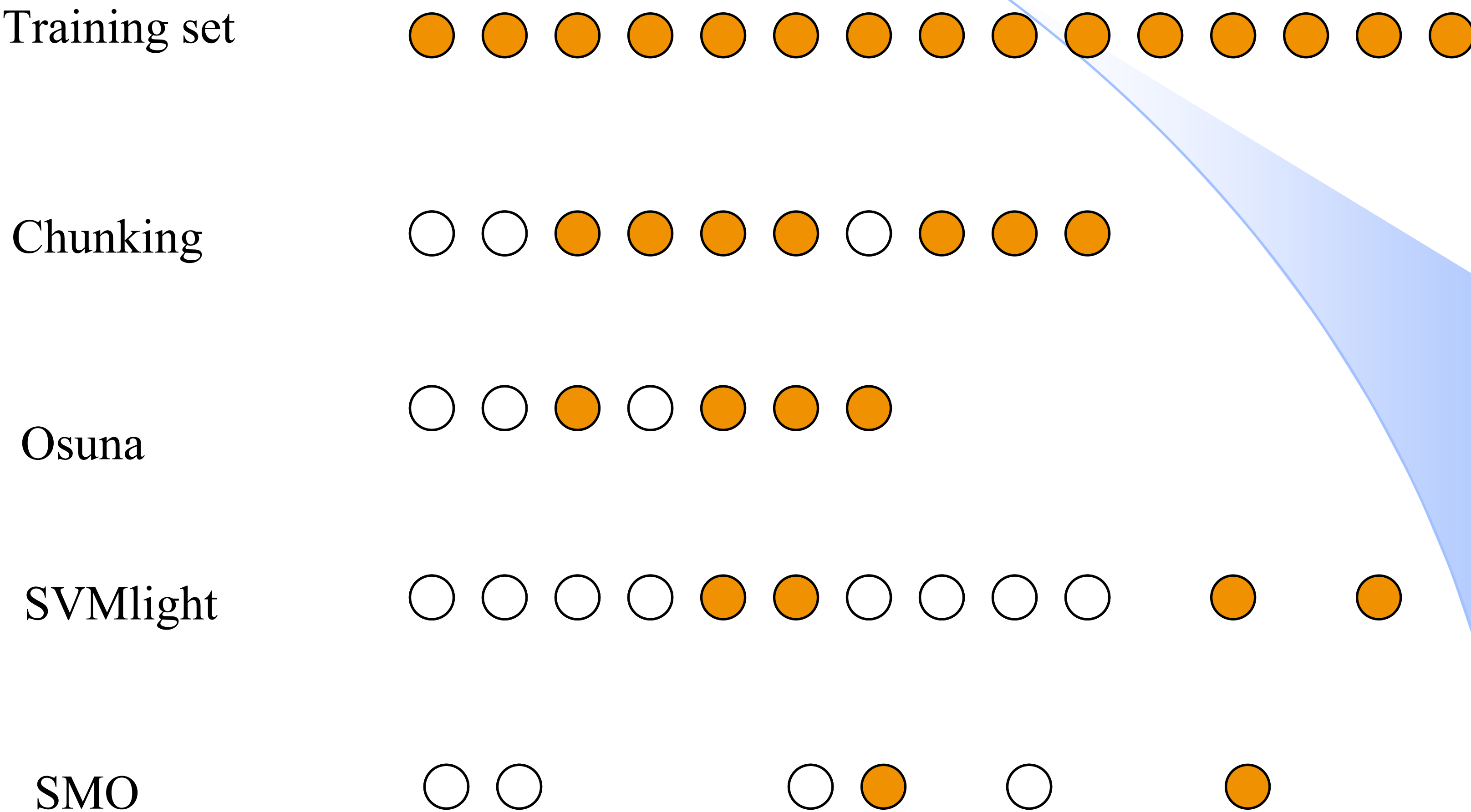
- Chuking
- Osuna
- SVMlight
- Sequential Minimal Optimization (SMO)

5. Implementation

Chunking



5. Implementation



5. Implementation

- Source code
  - SVM^{light}
  - LIBSVM
  - MySVM

6. Performance estimations

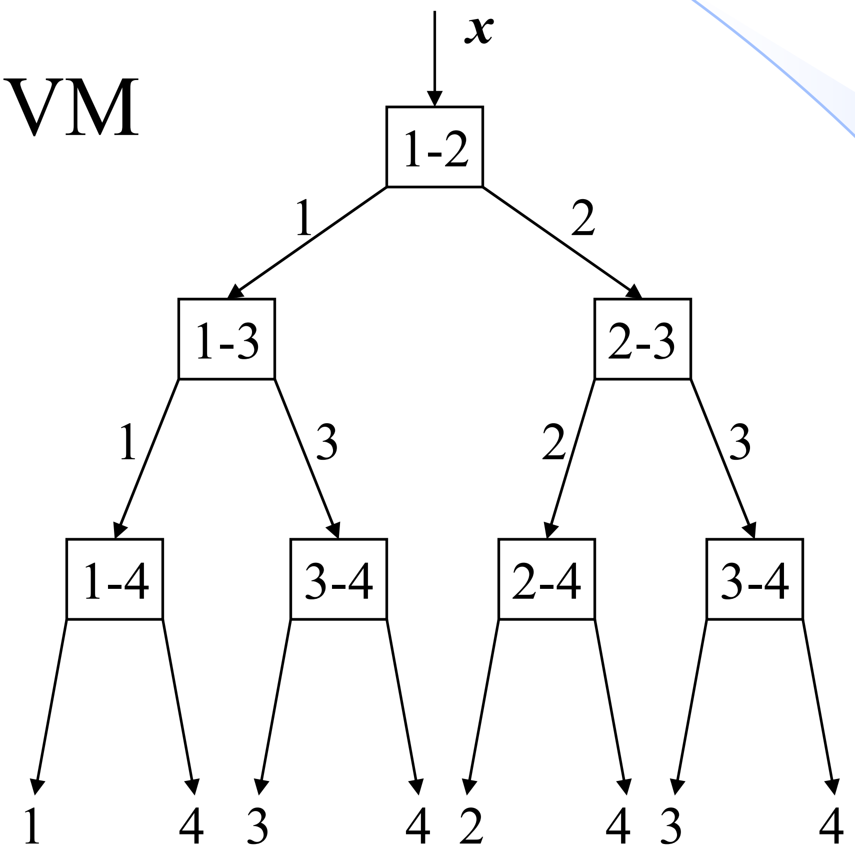
- Performances
  - Error rate
  - Recall
  - Precision
- Estimations
  - Cross Validation
  - Leave-one-out
  - $\xi\alpha - estimate$ (C-SVC)
  - $\xi\alpha\rho - estimate$ (v-SVC)
  - $\xi\alpha\rho\nu - estimate$ (One-class SVM)

7. Multi-class SVM

- One against all
- One against one
- DAG-SVM
- All together

7. Multi-class SVM

- DAG-SVM



8. Evolving SVM

- Objectives: $G\left(E_{\xi\alpha}, R_{\xi\alpha}, P_{\xi\alpha}\right)$
- Feature selection
- SVM training model
- Evolutionary Algorithms

Conclusions

- Efficiency ?
- Data set with noise ?
- So many classes ?
- Research and Implementation ?

THANK YOU!

Quang-Anh Tran

2003-2-28