Support Vector Machines and its Applications

Kris De Brabanter

Iowa State University

I. Support Vector Machines (SVM)

1. Linear SVM classifier: separable case
2. Linear SVM classifier: non-separable case
3. Nonlinear SVM classifier

II. Least Squares Support Vector Machines (LS-SVM)

4. Least Squares Support Vector Machines
   - Primal space (Classification)
   - Dual space (Classification)
   - Regression in dual space
   - Properties
   - Examples

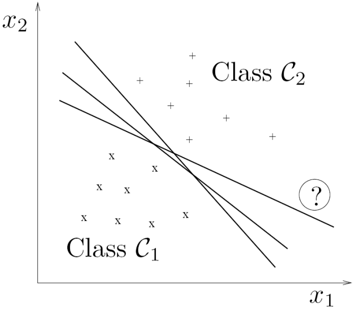
III. Fixed-Size Least Squares Support Vector Machines (FS-LSSVM)

5. Fixed-Size Least Squares Support Vector Machines
   - Problem formulation
   - Approximation for the feature map
   - Solution in primal space
   - Selection of support vectors
   - Examples and Comparison with SVM & LS-SVM

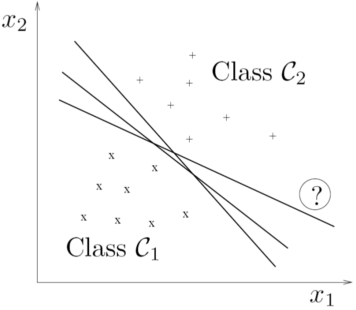
Part I

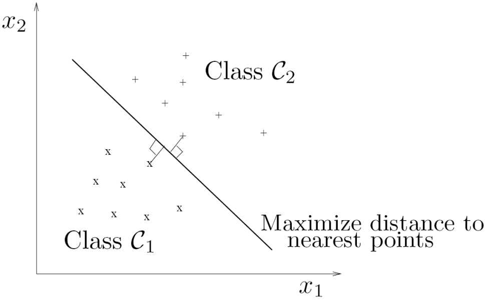
Support Vector Machines

Linear SVM classifier: separable case

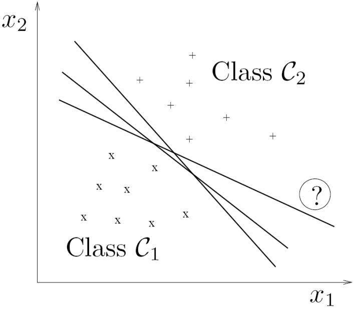


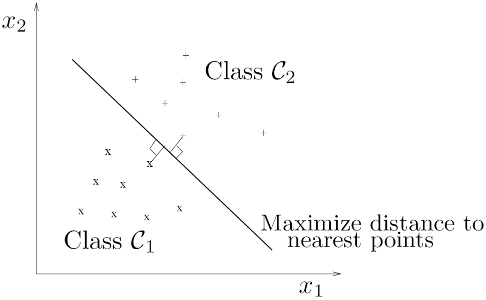
Linear SVM classifier: separable case





Linear SVM classifier: separable case

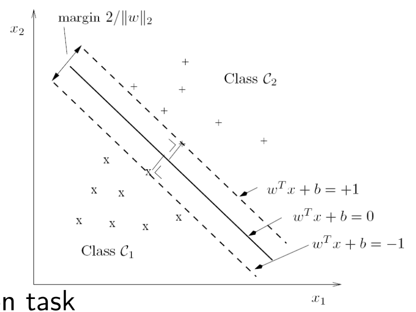




Separating hyperplane not unique

Unique hyperplane

Linear SVM classifier: separable case



- Classification task
- IID data $\mathcal{D} = \{(X_k, Y_k)\}_{i=1}^{n} \subset \mathbb{R}^d \times \{-1, +1\}$
- When data is separable

$$\begin{cases} w^T X_i + b \geq +1, & \text{if } Y_i = +1 \\ w^T X_i + b \leq -1 & \text{if } Y_i = -1. \end{cases}$$

Linear SVM classifier: separable case

Maximize the margin subject to the fact all training data need to be correctly classified (Vapnik & Lerner, 1963)

Linear SVM classifier: separable case

Maximize the margin subject to the fact all training data need to be correctly classified (Vapnik & Lerner, 1963)

Optimization problem (P)

$$\min_{w,b} \mathcal{J}_P(w) = \frac{1}{2} w^T w$$
$$s.t. \quad Y_k[w^T X_k + b] \geq 1, \quad k = 1, \ldots, n.$$

Linear SVM classifier: separable case

Maximize the margin subject to the fact all training data need to be correctly classified (Vapnik & Lerner, 1963)

Optimization problem (P)

$$\min_{w,b} \mathcal{J}_P(w) = \frac{1}{2} w^T w$$
$$s.t. \quad Y_k[w^T X_k + b] \geq 1, \quad k = 1, \ldots, n.$$

Using Lagrange multipliers α_k (dual problem)

Optimization problem (D)

$$\max_{\alpha} \mathcal{J}_D(\alpha) = -\frac{1}{2} \sum_{k,l=1}^{n} Y_k Y_l X_k^T X_l \alpha_k \alpha_l + \sum_{k=1}^{n} \alpha_k$$
$$s.t. \quad \sum_{k=1}^{n} \alpha_k Y_k = 0$$

Linear SVM classifier: separable case

1. Solution is given by

$$\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k X_k^T x + \hat{b}\right] \tag{1}$$

Linear SVM classifier: separable case

1. Solution is given by

$$\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k X_k^T x + \hat{b}\right] \tag{1}$$

2. Properties
   - Global and unique solution

Linear SVM classifier: separable case

1. Solution is given by

$$\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k X_k^T x + \hat{b}\right] \tag{1}$$

2. Properties
   - Global and unique solution
   - Sparseness $\Rightarrow$ many α_k are equal to zero

Linear SVM classifier: separable case

1. Solution is given by

$$\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k X_k^T x + \hat{b}\right] \tag{1}$$

2. Properties
   - Global and unique solution
   - Sparseness ⇒ many α_k are equal to zero
   - Eq (1) can be written as $\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{\sharp\mathrm{SV}} \hat{\alpha}_k Y_k X_k^T x + \hat{b}\right]$

Linear SVM classifier: separable case

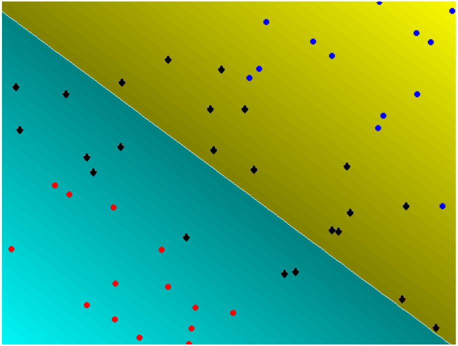
1. Solution is given by

$$\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k X_k^T x + \hat{b}\right] \tag{1}$$

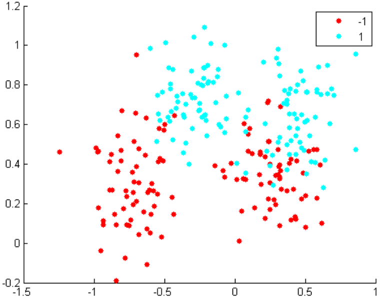
2. Properties
   - Global and unique solution
   - Sparseness $\Rightarrow$ many α_k are equal to zero
   - Eq (1) can be written as $\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{\sharp\mathrm{SV}} \hat{\alpha}_k Y_k X_k^T x + \hat{b}\right]$
   - Geometrical meaning of the support vectors

Linear SVM classifier: separable case

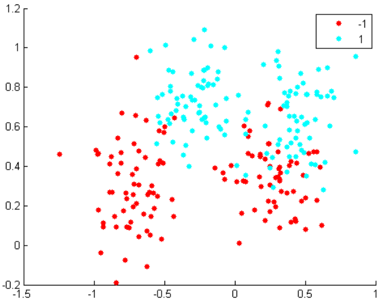
Toy Example: 25 data points in each class



Linear SVM classifier: non-separable case



Linear SVM classifier: non-separable case



- Tolerate misclassifications
- Introduce slack variables $\xi_k > 0$ (Cortes & Vapnik, 1995)
- $Y_k[w^T X_k + b] \geq 1 \rightarrow Y_k[w^T X_k + b] \geq 1 - \xi_k, \quad k = 1, \ldots, n$

Nonlinear SVM classifier

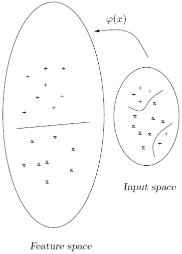
- Extension of linear classifier (non-separable case) to nonlinear classifier (Vapnik, 1995)

Nonlinear SVM classifier

- Extension of linear classifier (non-separable case) to nonlinear classifier (Vapnik, 1995)
- Input data mapped to a high dimensional feature space by nonlinear mapping φ

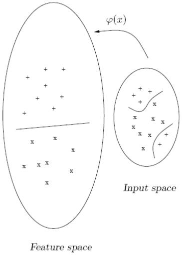
Nonlinear SVM classifier

- Extension of linear classifier (non-separable case) to nonlinear classifier (Vapnik, 1995)
- Input data mapped to a high dimensional feature space by nonlinear mapping φ



Nonlinear SVM classifier

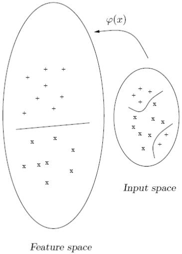
- Extension of linear classifier (non-separable case) to nonlinear classifier (Vapnik, 1995)
- Input data mapped to a high dimensional feature space by nonlinear mapping φ



- No explicit construction of φ needed (Mercer, 1909)

Nonlinear SVM classifier

- Extension of linear classifier (non-separable case) to nonlinear classifier (Vapnik, 1995)
- Input data mapped to a high dimensional feature space by nonlinear mapping φ



- No explicit construction of φ needed (Mercer, 1909)
- Using Mercer's condition: $K(x, z) = \varphi(x)^T \varphi(z)$ (Courant & Hilbert, 1953)

Nonlinear SVM classifier

Optimization problem (P)

$$\min_{w,b,\xi} \mathcal{J}_P(w,\xi) = \tfrac{1}{2} w^T w + c \textstyle\sum_{k=1}^{n} \xi_k$$

$$s.t. \quad Y_k[w^T \varphi(X_k) + b] \geq 1 - \xi_k, \quad k = 1, \ldots, n.$$

$$\xi_k \geq 0, \quad k = 1, \ldots, n.$$

Nonlinear SVM classifier

Optimization problem (P)

$$\min_{w,b,\xi} \mathcal{J}_P(w,\xi) = \tfrac{1}{2} w^T w + c \textstyle\sum_{k=1}^n \xi_k$$

$$\begin{aligned} s.t. \quad & Y_k[w^T \varphi(X_k) + b] \geq 1 - \xi_k, \quad k = 1,\ldots,n. \\ & \xi_k \geq 0, \quad k = 1,\ldots,n. \end{aligned}$$

Using Lagrange multipliers α_k (dual problem) + $K(X_k, X_l) = \varphi(X_k)^T \varphi(X_l)$

Optimization problem (D)

$$\max_{\alpha} \mathcal{J}_D(\alpha) = -\tfrac{1}{2} \textstyle\sum_{k,l=1}^n Y_k Y_l K(X_k, X_l) \alpha_k \alpha_l + \textstyle\sum_{k=1}^n \alpha_k$$

$$\begin{aligned} s.t. \quad & \textstyle\sum_{k=1}^n \alpha_k Y_k = 0 \\ & 0 \leq \alpha_k \leq c, \quad k = 1,\ldots,n. \end{aligned}$$

Nonlinear SVM classifier

1. Solution is given by

$$\hat{y}(x) = \text{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k K(X_k, x) + \hat{b}\right] \tag{2}$$

Nonlinear SVM classifier

1. Solution is given by

$$\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k K(X_k, x) + \hat{b}\right] \tag{2}$$

2. Properties
   - Global and unique solution

Nonlinear SVM classifier

1. Solution is given by

$$\hat{y}(x) = \text{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k K(X_k, x) + \hat{b}\right] \tag{2}$$

2. Properties
   - Global and unique solution
   - Sparseness $\Rightarrow$ many α_k are equal to zero

Nonlinear SVM classifier

1. Solution is given by

$$\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k K(X_k, x) + \hat{b}\right] \tag{2}$$

2. Properties
   - Global and unique solution
   - Sparseness $\Rightarrow$ many α_k are equal to zero
   - Eq (2) can be written as $\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{\sharp\mathrm{SV}} \hat{\alpha}_k Y_k K(X_k, x) + \hat{b}\right]$

Nonlinear SVM classifier

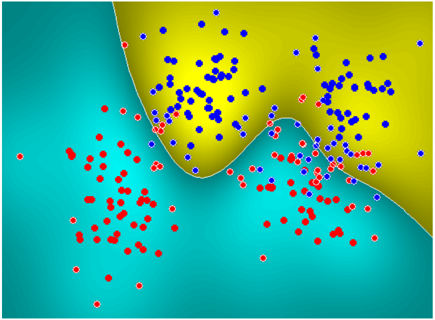
1. Solution is given by

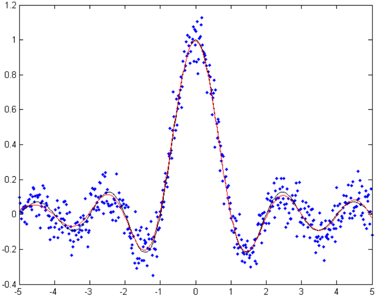
$$\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{n} \hat{\alpha}_k Y_k K(X_k, x) + \hat{b}\right] \tag{2}$$

2. Properties
   - Global and unique solution
   - Sparseness $\Rightarrow$ many α_k are equal to zero
   - Eq (2) can be written as $\hat{y}(x) = \operatorname{sign}\left[\sum_{k=1}^{\sharp\mathrm{SV}} \hat{\alpha}_k Y_k K(X_k, x) + \hat{b}\right]$
   - Geometrical meaning of the support vectors

nonlinear SVM classifier

1. LibSVM software (Chang & Lin, 2001)
2. Toy Example:
   - Ripley data set (250 data points)
   - Regression: sinc data (500 data points)





Part II

Least Squares Support Vector Machines

LS-SVM formulation

- Proposed by Suykens *et al.*, 1999

LS-SVM formulation

- Proposed by Suykens *et al.*, 1999
- WHY? Simplify the SVM formulation

LS-SVM formulation

- Proposed by Suykens *et al.*, 1999
- WHY? Simplify the SVM formulation
- Recall SVM formulation

Optimization problem (P)

$$\begin{aligned} \min_{w,b,\xi} \mathcal{J}_P(w,\xi) &= \tfrac{1}{2}w^T w + c\textstyle\sum_{k=1}^n \xi_k \\ s.t. \quad & Y_k[w^T\varphi(X_k)+b] \geq 1-\xi_k, \quad k=1,\ldots,n. \\ & \xi_k \geq 0, \quad k=1,\ldots,n. \end{aligned}$$

- LS-SVM formulation

Optimization problem (P)

$$\begin{aligned} \min_{w,b,e} \mathcal{J}_P(w,e) &= \tfrac{1}{2}w^T w + \tfrac{\gamma}{2}\textstyle\sum_{k=1}^n e_k^2 \\ s.t. \quad & Y_k[w^T\varphi(X_k)+b] = 1-e_k, \quad k=1,\ldots,n. \end{aligned}$$

LS-SVM formulation

- Using Lagrange multipliers, the solution is given by the LINEAR SYSTEM

$$\left(\begin{array}{c|c} 0 & Y^T \\ \hline Y & \Omega + \frac{1}{\gamma} I_n \end{array}\right) \left(\begin{array}{c} b \\ \hline \alpha \end{array}\right) = \left(\begin{array}{c} 0 \\ \hline 1_n \end{array}\right) \qquad (3)$$

- $\Omega = Y_k Y_l \varphi(X_k)^T \varphi(X_l) = Y_k Y_l K(X_k, X_l)$
- Classifier in dual space
  $\hat{y}(x) = \mathrm{sign}\left(\sum_{k=1}^{n} \hat{\alpha}_k Y_k K(x, X_k) + \hat{b}\right)$
- $K(\cdot, \cdot)$ has to be positive definite
- Extension to multiclass problems (Suykens *et al.*, 2002)

LS-SVM formulation for regression

- Analog to the classification case
- Using Lagrange multipliers, the solution is given by the LINEAR SYSTEM

$$\left(\begin{array}{c|c} 0 & 1_v^T \\ \hline 1_v & \Omega + \frac{1}{\gamma} I_n \end{array}\right) \left(\begin{array}{c} b \\ \hline \alpha \end{array}\right) = \left(\begin{array}{c} 0 \\ \hline Y \end{array}\right) \tag{4}$$

- $\Omega = \varphi(X_k)^T \varphi(X_l) = K(X_k, X_l)$
- Regressor in dual space $\hat{y}(x) = \sum_{k=1}^n \hat{\alpha}_k K(x, X_k) + \hat{b}$
- $K(\cdot, \cdot)$ has to be positive definite

Properties of the LS-SVM

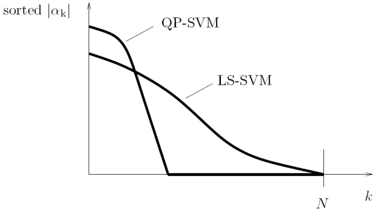
1. Advantages
   - Linear system instead of QP

Properties of the LS-SVM

1. Advantages
   - Linear system instead of QP
   - Global and unique solution

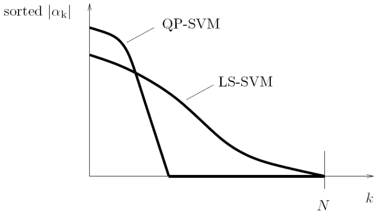
Properties of the LS-SVM

1. Advantages
    - Linear system instead of QP
    - Global and unique solution
2. Drawbacks
    - Lack of sparseness: $\alpha_k = \gamma e_k \Rightarrow$ Pruning techniques



Properties of the LS-SVM

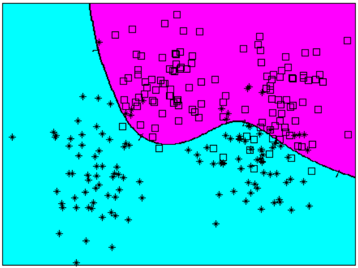
1. Advantages
    - Linear system instead of QP
    - Global and unique solution
2. Drawbacks
    - Lack of sparseness: $\alpha_k = \gamma e_k \Rightarrow$ Pruning techniques

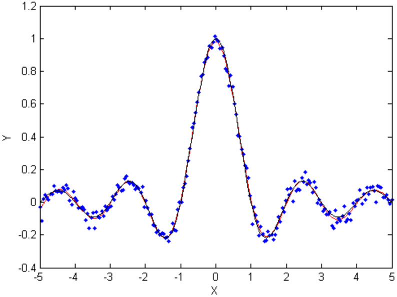


    - No geometrical interpretation of the support vectors

Examples

1. LS-SVMLab software (De Brabanter *et al.*, 2010)
2. Toy Example:
   - Ripley data set (250 data points)
   - Regression: sinc data (200 data points)





Part III

Fixed-Size Least Squares Support Vector Machines

Problem formulation

1. Can we solve the LS-SVM in primal space instead of dual?
2. Approximation of feature map φ needed
3. Is it possible to compute such a mapping?
4. What to do when data sets are large?
   - $N = 1000 \Rightarrow K \Rightarrow 8 \text{ MB}$
   - $N = 10000 \Rightarrow K \Rightarrow 763 \text{ MB}$
   - $N = 20000 \Rightarrow K \Rightarrow 3051 \text{ MB}$

Approximation for the feature map

(Nyström, 1930; Williams & Seeger, 2001)

- "big" matrix: $\Omega_{n,n} \in \mathbb{R}^{n \times n}$, "small" matrix: $\Omega_{m,m} \in \mathbb{R}^{m \times m}$ (based on e.g. random subsample, in practice often $m \ll n$)
- Eigenvalue decompositions: $\Omega_{n,n}\tilde{U} = \tilde{U}\tilde{\Lambda}$ and $\Omega_{m,m}\overline{U} = \overline{U}\,\overline{\Lambda}$
- Relation to eigenvalues and eigenfunctions of the integral equation

$$\int K(x,x')\phi_i(x)dF_X(x) = \lambda_i\phi_i(x')$$

with

$$\hat{\lambda}_i = \frac{1}{m}\overline{\lambda_i},\ \hat{\phi}_i(x_k) = \sqrt{m}\overline{u}_{ki},\ \hat{\phi}_i(x') = \frac{\sqrt{m}}{\overline{\lambda_i}}\sum_{k=1}^{m}\overline{u}_{ki}K(x_k,x')$$

Solution in primal space

- Feature map
  $\hat{\varphi}_i(x') = \sqrt{\lambda}_i \phi_i(x') = \frac{1}{\sqrt{\bar{\lambda}_i}} \sum_{k=1}^{m} \overline{u}_{ki} K(x_k, x'), i = 1, \ldots, m$

Solution in primal space

- Feature map
  $\hat{\varphi}_i(x') = \sqrt{\lambda}_i \phi_i(x') = \frac{1}{\sqrt{\lambda_i}} \sum_{k=1}^{m} \overline{u}_{ki} K(x_k, x'), i = 1, \ldots, m$
- recall

Optimization problem (P)

$$\min_{w,b} \mathcal{J}_P(w,b) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{n} \left(Y_k - (w^T \varphi(X_k) + b) \right)^2$$

- ridge regression in primal space (Suykens *et al.*, 2002)

Solution in primal space

- solution is given by

$$\begin{pmatrix} w \\ b \end{pmatrix} = \left(\hat{\Phi}_e^T \hat{\Phi}_e + \frac{I_{m+1}}{\gamma} \right)^{-1} \hat{\Phi}_e^T Y,$$

where $\hat{\Phi}_e$ is the $n \times (m+1)$ extended feature matrix

$$\hat{\Phi}_e = \begin{pmatrix} \hat{\varphi}_1(X_1) & \cdots & \hat{\varphi}_m(X_1) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \hat{\varphi}_1(X_n) & \cdots & \hat{\varphi}_m(X_n) & 1 \end{pmatrix}$$

- model has the form: $\hat{y}(x) = w^T \hat{\varphi}(x) + b$

Selection of support vectors

De Brabanter *et al.*, 2010

- Use entropy based criterion instead of random

Selection of support vectors

De Brabanter *et al.*, 2010

- Use entropy based criterion instead of random
- Quadratic Rényi entropy: $H_{R2}^{m}(x) = -\log \int f(x)^2 \, dx$

Selection of support vectors

De Brabanter *et al.*, 2010

- Use entropy based criterion instead of random
- Quadratic Rényi entropy: $H^m_{R2}(x) = -\log \int f(x)^2 \, dx$
- $H^m_{R2}(x) = -\log \frac{1}{m^2|H|^2} \sum_{k=1}^m \sum_{l=1}^m K\left\{\left(H\sqrt{2}\right)^{-1}(X_k - X_l)\right\}$

Selection of support vectors

De Brabanter *et al.*, 2010

- Use entropy based criterion instead of random
- Quadratic Rényi entropy: $H_{R2}^m(x) = -\log \int f(x)^2 \, dx$
- $H_{R2}^m(x) = -\log \frac{1}{m^2|H|^2} \sum_{k=1}^m \sum_{l=1}^m K\left\{\left(H\sqrt{2}\right)^{-1}(X_k - X_l)\right\}$
  - RBF Kernel: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$

Selection of support vectors

De Brabanter *et al.*, 2010

- Use entropy based criterion instead of random
- Quadratic Rényi entropy: $H^m_{R2}(x) = -\log \int f(x)^2\, dx$
- $H^m_{R2}(x) = -\log \frac{1}{m^2|H|^2} \sum_{k=1}^m \sum_{l=1}^m K\left\{\left(H\sqrt{2}\right)^{-1}(X_k - X_l)\right\}$
  - RBF Kernel: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$
  - H is called the bandwidth matrix, $H = \mathrm{diag}(h_1, \ldots, h_d)$

Selection of support vectors

De Brabanter *et al.*, 2010

- Use entropy based criterion instead of random
- Quadratic Rényi entropy: $H_{R2}^{m}(x) = -\log \int f(x)^2 \, dx$
- $H_{R2}^{m}(x) = -\log \frac{1}{m^2|H|^2} \sum_{k=1}^{m} \sum_{l=1}^{m} K\left\{ \left(H\sqrt{2}\right)^{-1} (X_k - X_l) \right\}$
  - RBF Kernel: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$
  - H is called the bandwidth matrix, $H = \mathrm{diag}(h_1, \ldots, h_d)$
  - Bandwidth choice $\rightarrow$ relation with density estimation: plug-in bandwidth selectors, solve-the-equation rules,...

Selection of support vectors

De Brabanter *et al.*, 2010

- Use entropy based criterion instead of random
- Quadratic Rényi entropy: $H_{R2}^{m}(x) = -\log \int f(x)^2 \, dx$
- $H_{R2}^{m}(x) = -\log \frac{1}{m^2|H|^2} \sum_{k=1}^{m} \sum_{l=1}^{m} K\left\{ \left(H\sqrt{2}\right)^{-1} (X_k - X_l) \right\}$
  - RBF Kernel: $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$
  - H is called the bandwidth matrix, $H = \mathrm{diag}(h_1, \ldots, h_d)$
  - Bandwidth choice $\rightarrow$ relation with density estimation: plug-in bandwidth selectors, solve-the-equation rules,...
  - For large data sets: fast evaluations of sums of Gaussians $\rightarrow$ Improved Fast Gauss Transform (Yang *et al.*, 2003), (Raykar & Duraiswami, 2006,2007)

Selection of support vectors

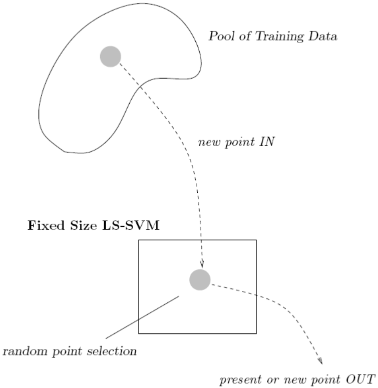
- Goal is to maximize this entropy criterion

Entropy Maximization

$$\begin{aligned} \max_f H_{R2}^m(f) &= -\log \int f^2(x)\,dx \\ s.t. \qquad & \int f(x)\,dx = 1 \\ & f(x) \geq 0. \end{aligned}$$

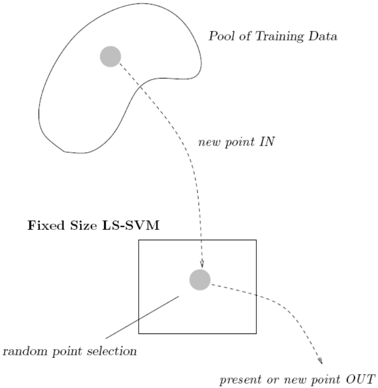
Selection of support vectors

- Maximizing the entropy: algorithm (Suykens *et al.*, 2002)



Selection of support vectors

- Maximizing the entropy: algorithm (Suykens *et al.*, 2002)

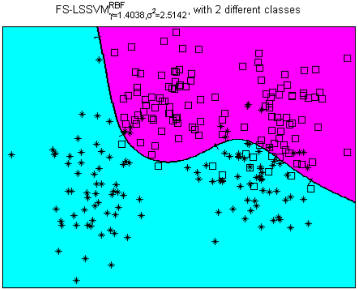


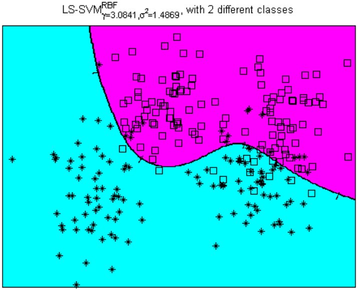
Examples

- Tuning parameters tuned with fast v-fold CV (De Brabanter *et al.*, 2010)
- Minimizing CV cost: CSA + gridsearch (Xavier de Souza *et al.*, 2006; De Brabanter *et al.*, 2010)
- Movie: regression, Motorcycle data set (133 data points)

Examples

- Tuning parameters tuned with fast v-fold CV (De Brabanter *et al.*, 2010)
- Minimizing CV cost: CSA + gridsearch (Xavier de Souza *et al.*, 2006; De Brabanter *et al.*, 2010)
- Movie: regression, Motorcycle data set (133 data points)
- Classification: Ripley data set (FS-LSSVM: 40 sv's)

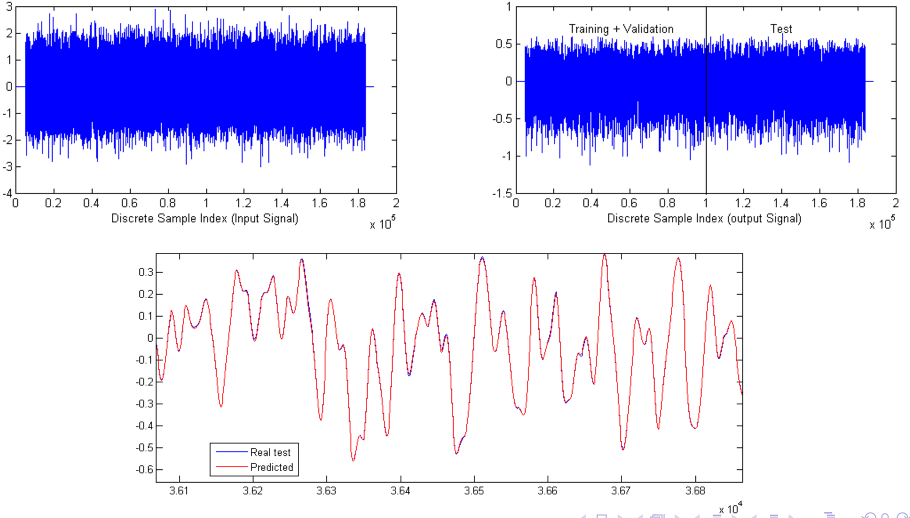




Examples

- System identification: large scale application (De Brabanter *et al.*, 2008b)
- SYSID2009 Wiener-Hammerstein Benchmark (188.000 data points, SISO system)
- Task: Given 100.000 training samples, simulate (iterative prediction) the following 88.000 samples
- Total of 2000 support vectors
- Selected kernel: RBF
- Tuning parameters tuned with fast v-fold CV
- Performance criteria: RMSE and fit percentage $f = 100\left(1 - \frac{\|y-\hat{y}\|}{\|y-\bar{y}\|}\right)$

Examples


3
2
1
0
-1
-2
-3
-4
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Discrete Sample Index (Input Signal)
x 10^5
1
0.5
0
-0.5
-1
-1.5
Training + Validation
Test
Discrete Sample Index (output Signal)
x 10^5
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
Real test
Predicted
3.61
3.62
3.63
3.64
3.65
3.66
3.67
3.68
x 10^4

Examples

- Results

Method	lags	$\text{RMSE}_{\text{test}}$	fit (%)
ARX	10	5.6×10^{-2}	76.47
MLP-NARX	11	2.3×10^{-2}	86.06
FS-LSSVM (Lin)	10	4.3×10^{-2}	81.93
FS-LSSVM (Poly)	10	6.0×10^{-3}	96.86
FS-LSSVM (RBF)	10	5.2×10^{-3}	97.78

Comparison with SVM & LS-SVM

- UCI data sets: Binary classification (misclassifications on test in %)

	spa	mgt	adu	ftc
N_{test}	1533	6020	12222	50000
n	57	11	14	54
♯ SV FS-LSSVM	200	1000	500	500
♯ SV C-SVC	800	7000	11085	185000
RBF FS-LSSVM	92.5(0.67)	86.6(0.51)	85.21(0.21)	81.8(0.52)
Lin FS-LSSVM	90.9(0.75)	77.8(0.23)	83.9(0.17)	75.61(0.35)
RBF C-SVC	92.6(0.76)	85.6(1.46)	84.81(0.20)	81.5(*)
Lin C-SVC	91.9(0.82)	77.3(0.53)	83.5(0.28)	75.24(*)
Maj. Rule	60.6(0.58)	65.8(0.28)	83.4(0.1)	51.23(0.20)

Comparison with SVM & LS-SVM

- UCI data sets: Binary classification (Computational time)

Av. Time (s)	spa	mgt	adu	ftc
RBF FS-LSSVM	44(5)	2103(64)	1601(208)	25160(523
Lin FS-LSSVM	15(0.8)	276(3.8)	304(12)	1114(15)
RBF C-SVC	1010(53)	20603(396)	139730(5556)	58962(*)
Lin C-SVC	785(22)	13901(189)	130590(4771)	53478(*)

(∗): no CV was performed. Timing is given for fixed tuning parameters

Comparison with SVM & LS-SVM

- UCI data sets: Regression

		bho	ccs
N_{test}		168	343
n		506	1030
$\sharp$ SV FS-LSSVM		135	120
$\sharp$ SV ε-SVR		226	670
RBF FS-LSSVM	L_2	0.13(0.02)	37.26(4.1)
	L_1	0.24(0.02)	4.47(0.26)
	L_∞	1.90(0.50)	27.78(5.43)
RBF ε-SVC	L_2	0.16(0.05)	62.24(5.8)
	L_1	0.24(0.03)	5.8(0.2)
	L_∞	2.20(0.54)	31.20(4.35)

Thanks for listening...

Questions???

References

Vapnik, V. and Lerner, A. (1963). Pattern Recogniction using Generalized Portrait Method, *Automation and Remote Control* 24:774–780.

Cortes, C. and Vapnik, V. (1995), Supoprt vector networks, *Machine Learning*, 20:273–297.

Vapnik, V. (1995), *The Nature of Statistical Learning*, Springer-Verlag, New York.

Mercer, J. (1909), Functions of positive and negative type and their connection with the theory of integral equations, *Philos. trans. Roy. Soc. London*, pp. 415–446.

Courant, R. and Hilbert, D. (1953), *Methods of Mathematical Physics*, International Science Publishers, New York.

Chang, C and Lin, C. (2001), LIBSVM: a library for support vector machines, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`

Suykens, J. A. K. and Vandewalle, J. (1999), Least squares support vector machine classifiers, *Neural Processing Letters*, 9(3): 293–300.

Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B. and Vandewalle, J. (2002), *Least Squares Support Vector Machines*. World Scientific, Singapore.

De Brabanter K., Karsmakers P., Ojeda F., Alzate C., De Brabanter J., Pelckmans K., De Moor B., Vandewalle J. and Suykens J.A.K., *LS-SVMlab Toolbox User's Guide version 1.8*, Internal Report 10-146, ESAT-SISTA, K.U.Leuven (Leuven, Belgium).

References

- Nyström, E. J. (1930), *Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben*, Acta Mathematica 54: 185–204.
- Williams, C. K. I. and Seeger, M. (2001), *Using the Nyström method to speed up kernel machines*, Advances in Neural Information Processing Systems.
- Girolami, M. (2002), *Orthogonal Series Density Estimation and the Kernel Eigenvalue Problem*, Neural Computation 14: 669-688.
- De Brabanter, K., De Brabanter, J., Suykens, J.A.K. and De Moor, B., *Optimized Fixed-Size Least Squares Support Vector Machines for Large Datasets*, Computational Statistics & Data Analysis, vol. 54, no. 6, Jun. 2010, pp. 1484-1504.
- Yang, C., Duraiswami, R., Gumerov, N. and Davis, L., (2003), *Improved fast Gauss transform and efficient kernel density estimation*. IEEE International Conference on Computer Vision 1, 464-471.
- Raykar, V. C. and Duraiswami, R., (2006). *Fast optimal bandwidth selection for kernel density estimation*. Proc. of the 2006 SIAM International Conference on Data Mining, Bethesda, Maryland.
- Raykar, V. C. and Duraiswami, R., (2007). Large-Scale Kernel Machines. MIT Press, Ch. The Improved Fast Gauss Transform with Applications to Machine Learning, pp. 175202.
- Xavier de Souza, S., Suykens, J. A. K., Vandewalle, J. and Bollé, D. (2006), *Cooperative Behavior in Coupled Simulated Annealing Processes with Variance Control*, Proc. of the International Symposium on Nonlinear Theory and its Applications (NOLTA2006), pp. 114-119.

References

De Brabanter K., Dreesen P., Karsmakers P., Pelckmans K., De Brabanter J., Suykens J.A.K. and De Moor B. (2008), *Fixed-Size LS-SVM Applied to the Wiener-Hammerstein Benchmark*, Internal Report 08-196, ESAT-SISTA, K.U.Leuven (Leuven, Belgium).