Least squares support vector machines

Prof. Dr. Ir. Bart De Moor

ESAT/SISTA
Katholieke Universiteit Leuven
Kasteelpark Arenberg 10
B-3001 Leuven
Belgium

T: +32(0)16321709
F: +32(0)16321970
E: bart.demoor@esat.kuleuven.ac.be
W: www.esat.kuleuven.ac.be/ demoor

Contents

- Least squares
  - Normal equations
  - Statistical interpretation
  - Regularized least squares
  - Trading-off bias versus variance
  - Weighted least squares
  - Weighted regularized least squares
  - Bayesian interpretation
- Least squares support vector machines
  - Intuitive idea
  - Mercer's condition
  - Least squares SVM
- Kernel PCA
- Kernel CCA
- Applications
- Research challenges

Linear Least Squares - Linear regression

Given A and b:

$$\min_x \|b - Ax\|_2^2 \,, \; A \in \mathbf{R}^{p \times q} \,, b \in \mathbf{R}^p \,.$$

Try to write b 'optimally' as a linear combination of the columns of A;
Let $f(x)$ be the objective function:

$$\begin{aligned} f(x) &= x^T A^T A x - 2b^T A x + b^T b \\ \frac{\partial f}{\partial x} &= 2A^T A x - 2A^T b = 0 \\ \frac{\partial^2 f}{\partial x} &= A^T A \end{aligned}$$

Hence local mimimum satisfying:

Normal equations:

$$A^T A x = A^T b \,.$$

$$\frac{\partial^2 f}{\partial x^2} = A^T A \geq 0$$

Least squares solutions x follow from

$$\boxed{A^T A x = A^T b}$$

Statistical interpretation of least squares

Let

$$A\hat{x} = \hat{b} \ ,$$

with $\mathrm{rank}(A) = q$ so that

$$\hat{x} = (A^T A)^{-1} A^T \hat{b} \ .$$

Let $\hat{b}$ be corrupted by additive noise f as

$$b = \hat{b} + f \ ,$$

with

$$\mathbf{E}(f) = 0 \ , \ \mathbf{E}(f f^T) = \sigma_f^2 I_p \ .$$

Least squares:

$$\min_x \|b - Ax\|_2^2 \Longrightarrow x = (A^T A)^{-1} A^T b \ .$$

Then

$$\mathbf{E}(x) = (A^T A)^{-1} A^T \mathbf{E}(b) = (A^T A)^{-1} A^T \hat{b} = \hat{x} \ .$$

Also

$$x - \hat{x} = (A^T A)^{-1} A^T f \ ,$$

so that the error-covariance matrix satisfies

$$\begin{aligned} \mathbf{E}[(x - \hat{x})(x - \hat{x})^T] &= (A^T A)^{-1} A^T \mathbf{E}(f f^T) A (A^T A)^{-1} \\ &= (A^T A)^{-1} \sigma_f^2 \ . \end{aligned}$$

Hence, the error covariance matrix is, up to within a scalar σ_f^2 given by $(A^T A)^{-1}$.

In particular, if A is ill-conditioned, in some directions, the error covariance matrix is large.

Regularized least squares

$$\min_{x\in\mathbf{R}^q} \|b - Ax\|_2^2 + \gamma x^T x$$

with $\gamma \in \mathbf{R}_0^+$ given regularization parameter.

Motivation: When A is ill-conditioned

- The norm of $x = A^\dagger b = V_1.S_1^{-1}.U_1^T b$ (as some singular values in S_1 are small) tends to be large, hence we append the square of norm of x to the objective function in an attempt to decrease it;
- $(A^T A)^{-1}$ is 'large': Large error covariance. Regularization will decrease it (to be shown);

Note that

* $\gamma \to \infty$, large importance to norm of x;
* $\gamma \to 0$, norm of x not important;

Solution:

$$\frac{\partial f}{\partial x} = 0 \Longrightarrow (A^T A + \gamma I_q)x = A^T b$$

Hence

$$x = (A^T A + \gamma I_q)^{-1} A^T b \ .$$

Note that

$$\lim_{\gamma \to 0} x(\gamma) = x_{\text{minimum norm}} = A^{\dagger} b \ .$$

Also note that

$$\lim_{\gamma \to \infty} x(\gamma) = 0 \ .$$

Trading-off bias versus variance

$$
\begin{aligned}
(A^T A)^{-1}\hat{x} &= A^T\hat{b} \\
(A^T A + \gamma I_q)x &= A^T b = A^T\hat{b} + A^T f
\end{aligned}
$$

Hence

$$
\begin{aligned}
x - \hat{x} &= -(I_q + (A^T A)/\gamma)^{-1}(A^T A)^{-1}A^T\hat{b} \\
&\quad +(A^T A + \gamma I_q)^{-1}A^T f \\
&\quad (\text{bias term}) + (\text{random term})
\end{aligned}
$$

Norm of bias squared:

$$\|\text{bias}\|^2 = \hat{x}^T(I_q + (A^T A)/\gamma)^{-2}\hat{x}\ .$$

Variance of random term

$$\text{trace}(A^T A + \gamma I_q)^{-1}(A^T A)(A^T A + \gamma I_q)^{-1}\sigma_f^2\ .$$

Note:

	$\|\text{bias}\|^2$	variance
$\gamma \to 0$	0	$\text{trace}((A^T A)^{-1}\sigma_f^2)$
$\gamma \to \infty$	$\|\hat{x}\|^2$	0

Variance is always smaller in PD sense

Is $\forall\gamma > 0$:

$(A^TA + \gamma I_q)^{-1}(A^TA)(A^TA + \gamma I_q)^{-1}\sigma_f^2 \leq (A^TA)^{-1}\sigma_f^2$?

Yes, use SVD of A and fact that

$$\sigma_i^2/(\sigma_i^2 + \gamma^2)^2 \leq 1/\sigma_i^2 \;, \forall i \;.$$

Condition number is always better

$$\kappa(A^TA + \gamma I_q) \leq \kappa(A^TA)$$

follows from

$$\frac{\sigma_1^2 + \gamma}{\sigma_q^2 + \gamma} \leq \frac{\sigma_1^2}{\sigma_q^2} \;,\; \gamma > 0.$$

Note that

$$\lim_{\gamma\to\infty} \kappa(A^TA + \gamma I_q) = 1 \text{ perfect condition}$$

Weighted least squares

Let $W = W^T$ be given (often but not always pos.def.)

$$\min_{x \in \mathbf{R}^q} \|b - Ax\|_W^2 = (b - Ax)^T W (b - Ax) \ .$$

Hence

$$(A^T W A)x = (A^T W b) \ .$$

If $\mathrm{rank}(A^T W A) = q$:

$$x = (A^T W A)^{-1} A^T W b \ .$$

Residuals:

$$\begin{aligned} e &= b - Ax = b - A(A^T W A)^{-1} A^T W b \\ &= (I_p - A(A^T W A)^{-1} A^T W))b \end{aligned}$$

Hence

$$A^T W e = 0 \ .$$

The residuals are 'W-orthogonal' to the column space of A.

Weighted and regularized least squares as a constrained optimization problem

Objective function for given W and V:

$$J = \frac{1}{2}e^T W^{-1} e + \frac{1}{2} x^T V^{-1} x \ ,$$

subject to

$$y = Ax + e \ .$$

Lagrangean:

$$\mathcal{L}(x, e, l) = J + l^T (y - Ax - e) \ .$$

This results in

$$\begin{aligned} W^{-1} e &= l \\ V^{-1} x &= A^T l \\ y &= Ax + e \end{aligned}$$

Three ways of solution:

<u>'Inner' solution in x: Eliminate e and l:</u>

$$x = (A^T W^{-1} A + V^{-1})^{-1} A^T W^{-1} y$$

<u>'Outer' solution in l: Eliminate e and x:</u>

$$l = (AVA^T + W)^{-1} y \; .$$

<u>'Outer' solution in e: Eliminate l and x:</u>

$$e = W(AVA^T + W)^{-1} y \; .$$

<u>Objective function:</u>

$$J = \frac{1}{2} y^T (AVA^T + W)^{-1} y \; .$$

Orthogonality of residuals and regressors

Let $W^{-1/2}$ and $V^{-1/2}$ be symmetric square roots of W. resp. V. Then

$$J = \frac{1}{2} \| \begin{pmatrix} W^{-1/2} y \\ 0 \end{pmatrix} - \begin{pmatrix} W^{-1/2} A x \\ V^{-1/2} x \end{pmatrix} \|_F^2 \; .$$

Obviously, the 'classical' least squares solution is

$$x = (A^T W^{-1} A + V^{-1}) A^T W^{-1} y \; .$$

Straightforward to prove:

$$\begin{aligned} & \begin{pmatrix} x^T A^T W^{-1/2} & x^T V^{-1/2} \end{pmatrix} \\ \times & \left[\begin{pmatrix} W^{-1/2} y \\ 0 \end{pmatrix} - \begin{pmatrix} W^{-1/2} A x \\ V^{-1/2} x \end{pmatrix} \right] \\ = & \; 0 \; . \end{aligned}$$

Bayesian interpretation

Model:

$$y = Az + f \ .$$

Statistical assumptions:

$$\begin{aligned} f &\sim \mathcal{N}(0, W) \qquad \mathbf{E}(zf^T) = 0 \\ z &\sim \mathcal{N}(0, V) \end{aligned}$$

Hence

$$\begin{aligned} y &\sim \mathcal{N}(0, (AVA^T + W)) \\ l &\sim \mathcal{N}(0, (AVA^T + W)^{-1}) \\ e &\sim \mathcal{N}(0, W(AVA^T + W)^{-1}W) \\ x &\sim \mathcal{N}(0, ...) \end{aligned}$$

Least squares support vector machines

- LS-SVM map a 'difficult' (nonlinear) regression problem in a low-dimensional space into an 'easy' (regularized) linear least squares problem in a high-dimensional (possibly infinite) space;

- Price to be paid: Computational complexity: Solve a set of linear equations of the size of the number of data points;

- LS-SVM can be used both for (nonlinear) classification (Example) as well as for nonlinear regression;

- Least-squares SVM are a least squares version of Vapnik's original SVM (which requires the solution of a QP problem the size of the number of data points).

Towards Mercer's condition

Consider two real variables $x, y \in \mathbf{R}$. Recall: McLaurin series of

$$
\begin{aligned}
e^{-x^2} &= 1 - x^2 + 1/2\,x^4 - 1/6\,x^6 + 1/24\,x^8 - \frac{1}{120}\,x^{10} \\
&\quad + \frac{1}{720}\,x^{12} - \frac{1}{5040}\,x^{14} + \ldots
\end{aligned}
$$

Also

$$
\begin{aligned}
e^{-\frac{(x-y)^2}{\sigma^2}} &= 1 - \frac{(\mathtt{x}-y)^2}{\sigma^2} + 1/2\,\frac{(\mathtt{x}-y)^4}{\sigma^4} - 1/6\,\frac{(\mathtt{x}-y)^6}{\sigma^6} \\
&\quad + 1/24\,\frac{(\mathtt{x}-y)^8}{\sigma^8} - \frac{1}{120}\,\frac{(\mathtt{x}-y)^{10}}{\sigma^{10}} \\
&\quad + \frac{1}{720}\,\frac{(\mathtt{x}-y)^{12}}{\sigma^{12}} - \frac{1}{5040}\,\frac{(\mathtt{x}-y)^{14}}{\sigma^{14}} \\
&\quad + \frac{1}{40320}\,\frac{(\mathtt{x}-y)^{16}}{\sigma^{16}} - \frac{1}{362880}\,\frac{(\mathtt{x}-y)^{18}}{\sigma^{18}} \\
&\quad + \frac{1}{3628800}\,\frac{(\mathtt{x}-y)^{20}}{\sigma^{20}} - \frac{1}{39916800}\,\frac{(\mathtt{x}-y)^{22}}{\sigma^{22}} \\
&\quad + \frac{1}{479001600}\,\frac{(\mathtt{x}-y)^{24}}{\sigma^{24}} + \ldots
\end{aligned}
$$

It can be seen that this expression is equal to

$$
\begin{aligned}
e^{-\frac{(x-y)^2}{\sigma^2}} &= \begin{pmatrix} 1 & x & x^2 & x^3 & x^4 & \ldots \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \ldots \\ 0 & \frac{1}{\sigma} & 0 & \ldots \\ 0 & 0 & \frac{1}{\sigma^2} & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \\
&\times \begin{pmatrix}
1 & 0 & -1 & 0 & \frac{1}{2} & 0 & -\frac{1}{6} & 0 & \frac{1}{24} & \ldots \\
0 & 2 & 0 & -\frac{4}{2} & 0 & \frac{6}{6} & 0 & -\frac{8}{24} & 0 & \ldots \\
-1 & 0 & \frac{6}{2} & 0 & -\frac{15}{6} & 0 & \frac{28}{24} & 0 & -\frac{45}{120} & \ldots \\
0 & -\frac{4}{2} & 0 & \frac{20}{6} & 0 & -\frac{56}{23} & 0 & \frac{120}{120} & 0 & \ldots \\
\frac{1}{2} & 0 & -\frac{15}{6} & 0 & \frac{70}{24} & 0 & -\frac{210}{120} & 0 & \frac{495}{720} & \ldots \\
0 & \frac{6}{6} & 0 & -\frac{56}{24} & 0 & \frac{252}{120} & 0 & -\frac{792}{720} & 0 & \ldots \\
-\frac{1}{6} & 0 & \frac{28}{24} & 0 & -\frac{210}{120} & 0 & \frac{924}{720} & 0 & -\frac{3003}{5040} & \ldots \\
0 & -\frac{8}{24} & 0 & \frac{120}{120} & 0 & -\frac{792}{720} & 0 & \frac{3432}{5040} & 0 & \ldots \\
\frac{1}{24} & 0 & -\frac{45}{120} & 0 & \frac{495}{720} & 0 & -\frac{3003}{5040} & 0 & \frac{12870}{40320} & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ldots
\end{pmatrix} \\
&\times \begin{pmatrix} 1 & 0 & 0 & \ldots \\ 0 & \frac{1}{\sigma} & 0 & \ldots \\ 0 & 0 & \frac{1}{\sigma^2} & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} 1 \\ y \\ y^2 \\ y^3 \\ y^4 \\ \vdots \end{pmatrix} .
\end{aligned}
$$

It can be shown that the middle matrix is **positive definite** as follows:

$$
\begin{aligned}
e^{-(x-y)^2} &= e^{-x^2}\ e^{2xy}\ e^{-y^2} \\
&= e^{-x^2}(1 + 2xy + \frac{2^2}{2!}x^2y^2 + \frac{2^3}{3!}x^3y^3 + \ldots)e^{-y^2}
\end{aligned}
$$

$$= e^{-x^2}(1\ x\ x^2\ x^3\ \ldots)\begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & \ldots \\ 0 & 2 & 0 & 0 & \ldots & 0 & \ldots \\ 0 & 0 & \frac{2^2}{2!} & 0 & \ldots & 0 & \ldots \\ 0 & 0 & 0 & \frac{2^3}{3!} & \ldots & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \frac{2^n}{n!} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ y \\ y^2 \\ \vdots \end{pmatrix} e^{-y^2}$$

Define the vector function

$$\phi^T(x) = e^{-x^2}(1\ x\ x^2\ x^3\ \ldots)\mathrm{diag}(\sqrt{\frac{2^n}{n!}})\ .$$

Then

$$\boxed{e^{-(x-y)^2} = \phi^T(x).\phi(y)}$$

This is an inner product of two infinite dimensional vectors !

Mercer's condition

If $u_i \in \mathbf{R}^n, i = 1, \ldots, N$, then the matrix Ω with elements

$$\Omega_{ij} = e^{-\frac{||u_i - u_j||^2}{\sigma^2}}$$

is positive definite.

Functions of the form

$$e^{-\frac{||u_i - u_j||^2}{\sigma^2}}$$

are called Radial Basis Functions.

$\Longrightarrow$ Close relation between (LS-)SVM and RBF-Neural Nets

Least squares support vector machine

Consider the static input-output model

$$y = Ax + e \ .$$

with $y \in \mathbf{R}^N$ containing N consecutive 'outputs', and

$$A = \begin{pmatrix} \phi^T(u_1) \\ \phi^T(u_2) \\ \vdots \\ \phi^T(u_N) \end{pmatrix} \ .$$

The vector $x \in \mathbf{R}^{\infty \times n}$ is an infinite vector of unknowns.

The regularized linear regression problem to estimate x is:

$$\min_{x,e} \ x^T x + \frac{1}{\gamma} \, e^T e$$

subject to

$$y = Ax + e \ .$$

The positive scalar $\gamma \in \mathbf{R}_0^+$ makes the trade off between the norm of x and the norm of e.

Solution

As seen before, with the vector of Lagrange multipliers $l \in \mathbf{R}^N$:

$$\begin{aligned} x &= A^T l \\ e &= l\gamma \\ y &= Ax + e \end{aligned}$$

results in

$$y = (AA^T + \gamma I_N)l \ ,$$

Because of Mercer's condition, AA^T is a finite, $N \times N$ positive definite matrix.

$$l = (AA^T + \gamma I_N)^{-1} y \ .$$

Remarks:

- The solution for x would be infinite dimensional, but for l it is finite dimensional !
- There are only two 'design' parameters:

  σ: Variance of the Radial Basis Function

  γ Regularization parameter; As $\gamma \infty 0 \Longrightarrow$ perfect fit !

Note that

$$A.A^T = \begin{pmatrix} \cdots & \vdots & \cdots \\ \cdots & e^{-\|u_i - u_j\|^2/\sigma^2} & \cdots \\ \cdots & \vdots & \cdots \end{pmatrix}$$

is a finite $N \times N$ **positive definite** matrix.

Some demo's

```
d:\matlab\lssvm\demo.m
d:\matlab\lssvm\demo2.m
```

Complexity reduction via PCA

Let $A \in \mathbf{R}^{p \times q}$ be a given data matrix of full column rank $r_A = q$ with SVD (dyadic decomposition):

$$A = U \Sigma V^T = \sum_{i=1}^{q} u_i . \sigma_i . v_i^T \ .$$

The sample covariance matrix is defined as AA^T/q. From SVD:

$$AA^T = (U\Sigma V^T)(V\Sigma^T U^T) = U(\Sigma\Sigma^T)U^T = \sum_{i=1}^{p} u_i \sigma_i^2 u_i^T \ .$$

Hence, the dyadic decomposition of AA^T

- is a sum of symmetric rank one matrices, of decreasing variance σ_i^2;
- can be calculated from the singular values and left singular vectors of A;

Principal component analysis

Let $A \in \mathbf{R}^{p\times q}$ be a given data matrix.

PCA

$=$ dimensionality reduction on the covariance matrix AA^T

$=$ least squares approxiation of AA^T by covariance matrix BB^T of lower rank

$=$ truncate the dyadic decomposition of AA^T

$=$ B also obtained from truncating dyadic decomposition of A;

Kernel PCA

Given data vectors $x_i \in \mathbf{R}^n, i = 1, \ldots, N$.

Consider the matrix of transformed data:

$$A = \begin{pmatrix} \phi^T(x_1) \\ \phi^T(x_2) \\ \vdots \\ \phi^T(x_N) \end{pmatrix} .$$

and its covariance matrix

$$\Omega_{ij} = (AA^T)_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}} .$$

Apply PCA (EVD) to Ω !

Canonical correlation analysis

Just as PCA is the 'covariance' formulation of the Eckart-Young-Mirksy Theorem, CCA is the 'covariance' formulation of principal angles and directions.
Given two data matrices A and B;

Consider the sample covariance matrix

$$\begin{pmatrix} A^T \\ B^T \end{pmatrix} . \begin{pmatrix} A & B \end{pmatrix} = \begin{pmatrix} A^T A & A^T B \\ B^T A & B^T B \end{pmatrix}$$

The eigenvale problem

$$\begin{pmatrix} 0 & A^T B \\ B^T A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} A^T A & 0 \\ 0 & B^T B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \lambda$$

with

$$\begin{pmatrix} x^T & y^T \end{pmatrix} \begin{pmatrix} A^T A & 0 \\ 0 & B^T B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1 \; .$$

delivers

The principal angles: $\lambda_i = \pm \cos \theta_i$;

The principal directions: Ax and By;

Kernel CCA

Two data sets $x_i \in \mathbf{R}^n, i = 1, \ldots, M$ and $y_j \in \mathbf{R}^n, j = 1, \ldots, N$;

Consider the matrices

$$A = \begin{pmatrix} \phi^T(x_1) \\ \phi^T(x_2) \\ \vdots \\ \phi^T(x_M) \end{pmatrix} \; ; \; B = \begin{pmatrix} \phi^T(y_1) \\ \phi^T(y_2) \\ \vdots \\ \phi^T(y_N) \end{pmatrix} \; .$$

Apply CCA to

$$\begin{pmatrix} A^T \\ B^T \end{pmatrix} . \begin{pmatrix} A & B \end{pmatrix} = \begin{pmatrix} A^T A & A^T B \\ B^T A & B^T B \end{pmatrix}$$

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ \ldots & e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}} & \ldots & e^{-\frac{\|x_i - y_j\|^2}{\sigma^2}} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \ldots & e^{-\frac{\|y_i - x_j\|^2}{\sigma^2}} & \ldots & e^{-\frac{\|y_i - y_j\|^2}{\sigma^2}} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Future research: Linear-nonlinear/static-dynamic

	Static	Dynamic
Linear	LS TLS PCA CCA	PEM STLS SS CVA
Kernel	LS-SVM TLS-SVM ? Kernel PCA ? Kernel CCA ?	? ? ? ?

- Choice of regularization parameter
  - Multi-cross validation
  - Fitting versus generalization (prediction)
  - Bayesian (levels of inference)
- Solving large sets of linear equations
  - Methods from numerical linear algebra
  - Sparse approximations
- Input selection
  - Inherently combinatorial

  - Chicken-egg situation: Input selection versus modelling
  - Automatic Relevance Detection

- Applications
  - Financial engineering (time-series prediction)
  - Bio-informatics (gene expression modelling and classification)
  - Advanced Process Control (APC) (modelling and control, softsensors)
  - . . .

Some references

- Suykens J.A.K., Vandewalle J., "Least squares support vector machine classifiers", Neural Processing Letters, vol. 9, no. 3, Jun. 1999, pp. 293-300.
- Suykens J.A.K., Vandewalle J., De Moor B., "Optimal control by least squares support vector machines", Neural Networks, vol. 14, no. 1, Jan. 2001, pp. 23-35.
- Suykens J.A.K., Vandewalle J., "Recurrent least squares support vector machines", IEEE Transactions on Circuits and Systems-I, vol. 47, no. 7, Jul. 2000, pp. 1109-1114.
- Suykens J.A.K., Lukas L., Van Dooren P., De Moor B., Vandewalle J., "Least squares support vector machine classifiers : a large scale algorithm", in Proc. of the European Conference on Circuit Theory and Design (ECCTD'99), Stresa, Italy, Sep. 1999, pp. 839-842.
- Suykens J.A.K., "Least squares support vector machines for classification and nonlinear modelling", Neural Network World, Special Issue on PASE 2000, vol. 10, no. 1-2, Jan. 2000, pp. 29-48.

- Van Gestel T., Suykens J., Baesens B., Viaene S., Vanthienen J., Dedene G., De Moor B., Vandewalle J., "Benchmarking Least Squares Support Vector Machine Classifiers", Internal Report 00-37, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2000.
- Van Gestel T., Suykens J., Baestaens D., Lambrechts A., Lanckriet G., Vandaele B., De Moor B., Vandewalle J., "Financial Time Series Prediction using Least Squares Support Vector Machines within the Evidence Framework", Internal Report 00-79, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2000. Accepted for publication in IEEE Transactions on Neural Networks.