Least Squares Support Vector Regression with Applications to Large-Scale Data: a Statistical Approach

Kris De Brabanter

Public Defense

April, 27 2011

Promotor: Prof. dr. ir. B. De Moor
Co-Promotor: Prof. dr. ir. J. Suykens

Outline

1. Goal & Overview
2. Introduction
   - Parametric vs. nonparametric regression
   - Nonparametric regression estimates: an overview
3. Fixed-Size Least Squares Support Vector Machines
   - Fixed Size LS-SVM formulation
   - Selection of Support Vectors
   - Practical identification problem
4. Robust Nonparametric Methods
   - Problems with outliers
   - Robust nonparametric regression
5. Correlated Errors
   - Problems with correlation in nonparametric regression
   - Removing correlation effects
6. Confidence Intervals
7. Conclusions

Outline

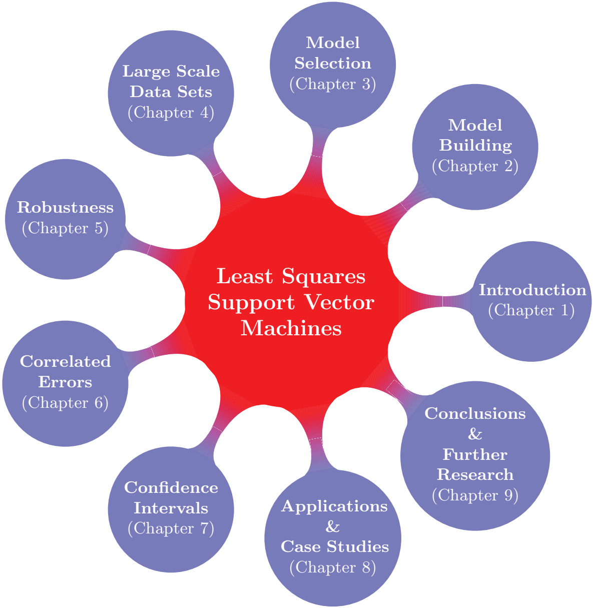
1. Goal & Overview
2. Introduction
   - Parametric vs. nonparametric regression
   - Nonparametric regression estimates: an overview
3. Fixed-Size Least Squares Support Vector Machines
   - Fixed Size LS-SVM formulation
   - Selection of Support Vectors
   - Practical identification problem
4. Robust Nonparametric Methods
   - Problems with outliers
   - Robust nonparametric regression
5. Correlated Errors
   - Problems with correlation in nonparametric regression
   - Removing correlation effects
6. Confidence Intervals
7. Conclusions

Goal of the Thesis

Goal of the Thesis

Study the properties of Least Squares Support Vector Machines for regression with an emphasis on statistical aspects and develop a framework for large scale data

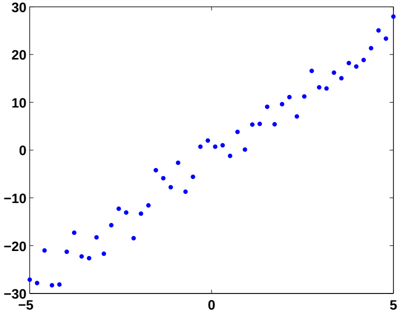
Overview


Least Squares Support Vector Machines
Introduction (Chapter 1)
Model Building (Chapter 2)
Model Selection (Chapter 3)
Large Scale Data Sets (Chapter 4)
Robustness (Chapter 5)
Correlated Errors (Chapter 6)
Confidence Intervals (Chapter 7)
Applications & Case Studies (Chapter 8)
Conclusions & Further Research (Chapter 9)

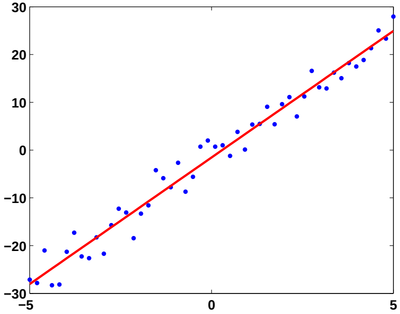
Outline

1. Goal & Overview
2. Introduction
   - Parametric vs. nonparametric regression
   - Nonparametric regression estimates: an overview
3. Fixed-Size Least Squares Support Vector Machines
   - Fixed Size LS-SVM formulation
   - Selection of Support Vectors
   - Practical identification problem
4. Robust Nonparametric Methods
   - Problems with outliers
   - Robust nonparametric regression
5. Correlated Errors
   - Problems with correlation in nonparametric regression
   - Removing correlation effects
6. Confidence Intervals
7. Conclusions

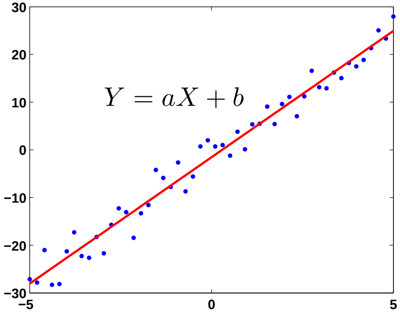
A simple example



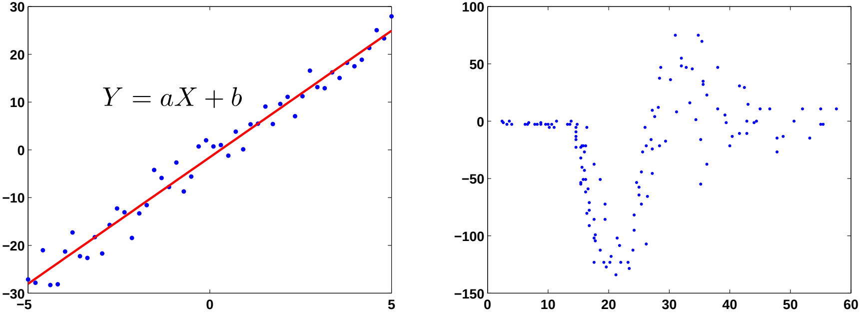
A simple example



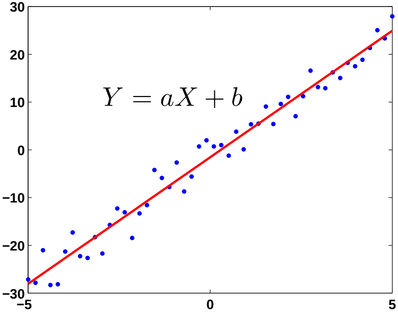
A simple example

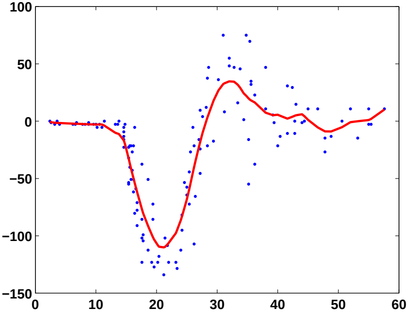


A simple example

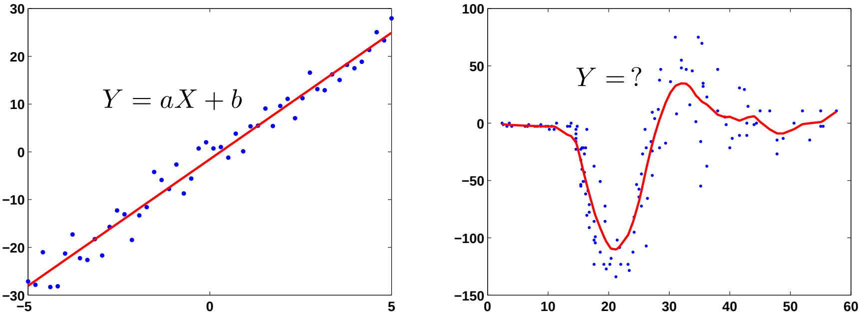


A simple example

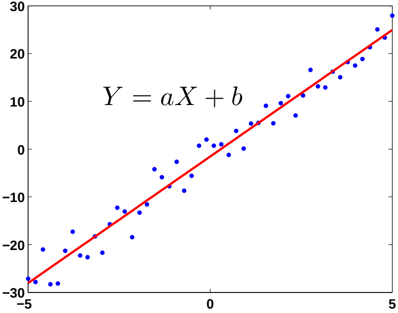


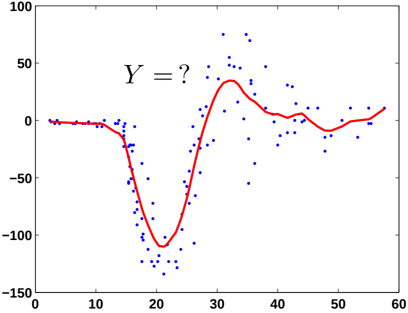


A simple example



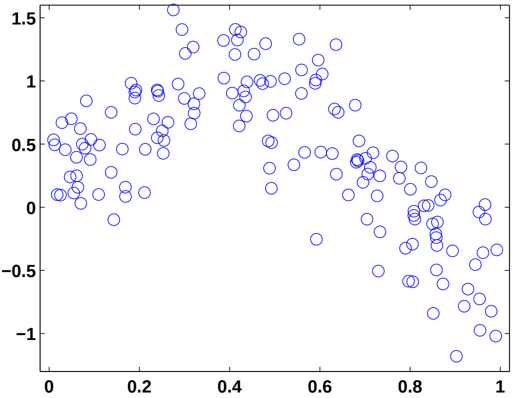
A simple example



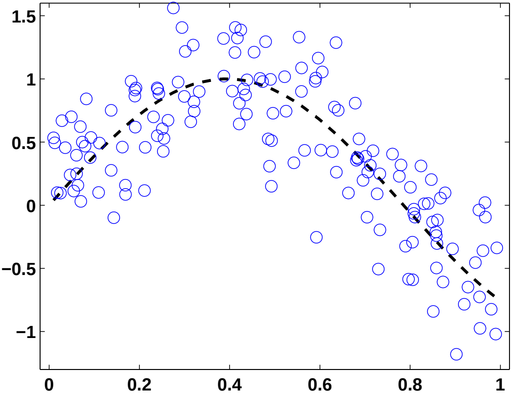


PARAMETRIC FORM IS NOT ALWAYS EASY TO FIND

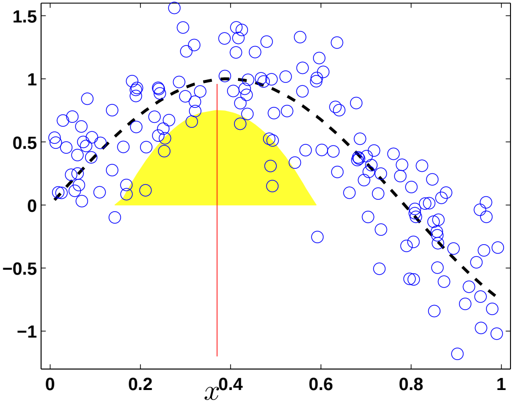
Construction of a nonparametric estimate: NW smoother



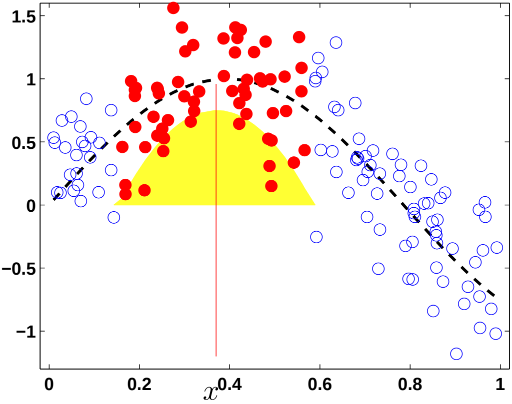
Construction of a nonparametric estimate: NW smoother



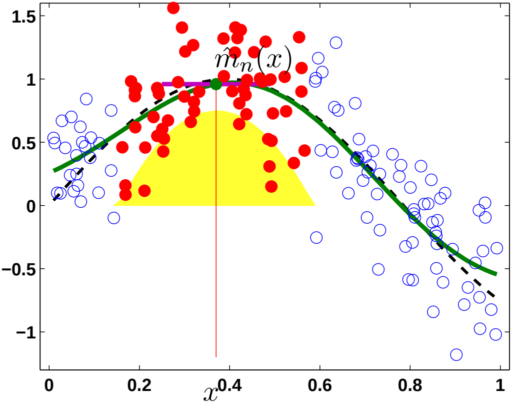
Construction of a nonparametric estimate: NW smoother



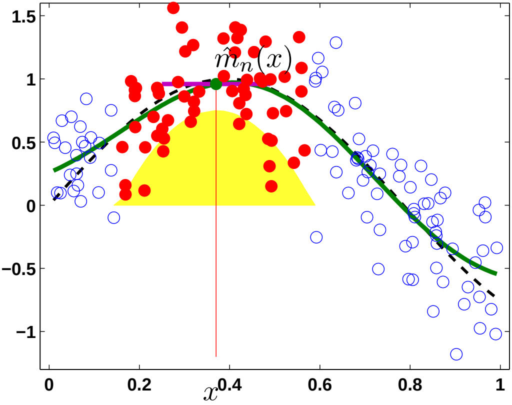
Construction of a nonparametric estimate: NW smoother



Construction of a nonparametric estimate: NW smoother



Construction of a nonparametric estimate: NW smoother



$$\hat{m}_n(x) = \sum_{i=1}^{n} \frac{K(\frac{x-X_i}{h})}{\sum_{j=1}^{n} K(\frac{x-X_j}{h})} Y_i$$

Other nonparametric regression estimates

- Local constant regression (Nadaraya, 1964; Watson, 1964)
- Regression trees (Breiman *et al.*, 1984)
- Wavelets (Daubechies, 1992)
- Nearest Neighbors (Devroye *et al.*, 1994)
- Local linear regression (Fan & Gijbels, 1996)
- Support vector machines (Vapnik, 1995)
- Splines (Wahba, 1990; Eubank, 1999)
- Partitioning estimates (Györfi *et al.*, 2002)
- **Least squares support vector machines** (Suykens *et al.*, 2002)
- ...

Least squares support vector machines

Primal formulation (LS-SVM formulation for regression)

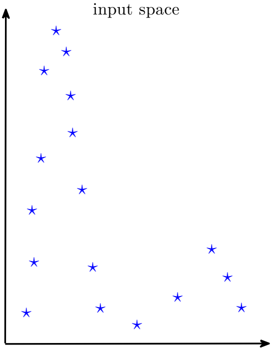
$$\min_{w,b,e} \mathcal{J}_P(w,e) = \tfrac{1}{2} w^T w + \tfrac{\gamma}{2} \textstyle\sum_{k=1}^{n} e_k^2$$
$$s.t. \quad w^T \varphi(X_k) + b + e_k = Y_k, \quad k = 1, \ldots, n.$$

Least squares support vector machines

Primal formulation (LS-SVM formulation for regression)

$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{k=1}^{n} e_k^2$$

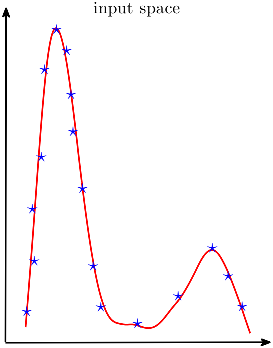
$$s.t. \quad w^T\varphi(X_k) + b + e_k = Y_k, \quad k = 1, \ldots, n.$$



Least squares support vector machines

Primal formulation (LS-SVM formulation for regression)

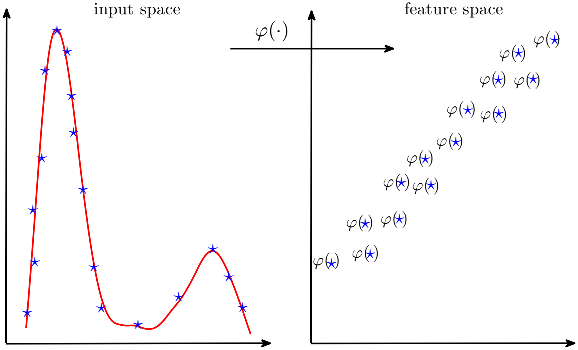
$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{k=1}^{n} e_k^2$$
$$s.t. \quad w^T \varphi(X_k) + b + e_k = Y_k, \quad k = 1, \ldots, n.$$



Least squares support vector machines

Primal formulation (LS-SVM formulation for regression)

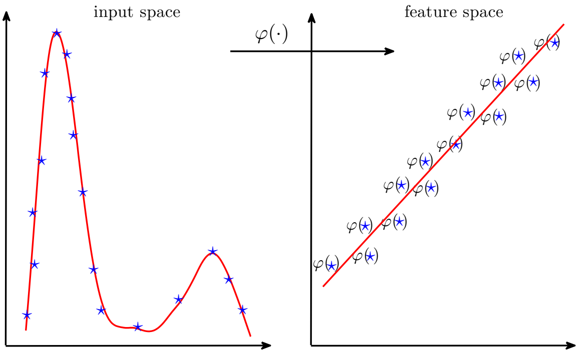
$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{k=1}^{n} e_k^2$$
$$s.t. \quad w^T \varphi(X_k) + b + e_k = Y_k, \quad k = 1, \ldots, n.$$



Least squares support vector machines

Primal formulation (LS-SVM formulation for regression)

$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{n} e_k^2$$
$$s.t. \quad w^T \varphi(X_k) + b + e_k = Y_k, \quad k = 1, \ldots, n.$$



LS-SVM: solution + model selection

- $\mathcal{D}_n = \{(X_k, Y_k) : X_k \in \mathbb{R}^d, Y_k \in \mathbb{R}; k = 1, \ldots, n\} \overset{\text{i.i.d.}}{\sim} (X, Y)$

Primal formulation

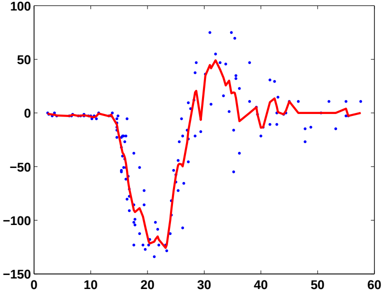
$$\min_{w,b,e} \mathcal{J}_P(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{n} e_k^2$$
$$s.t. \quad Y_k = w^T \varphi(X_k) + b + e_k, \quad k = 1, \ldots, n.$$

Dual formulation

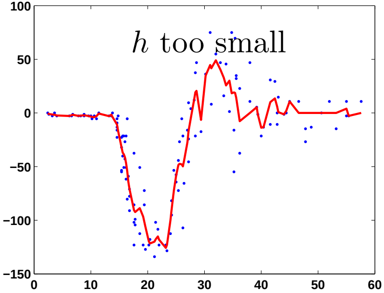
$$\left(\begin{array}{c|c} 0 & 1_n^T \\ \hline 1_n & \Omega + \frac{I_n}{\gamma} \end{array}\right) \left(\begin{array}{c} b \\ \hline \alpha \end{array}\right) = \left(\begin{array}{c} 0 \\ \hline Y \end{array}\right)$$

- $\Omega_{kl} = \varphi(X_k)^T \varphi(X_l) = K(X_k, X_l) = (2\pi)^{-d/2} \exp(-\frac{\|X_k - X_l\|^2}{2\mathbf{h}^2})$
- K has to be positive definite i.e. $\int \exp(-j\omega x) K(x)\, dx \geq 0$
- Model in dual space $\hat{m}_n(x) = \sum_{k=1}^{n} \hat{\alpha}_k K(x, X_k) + \hat{b}$
- γ and $\mathbf{h}$: tuning parameters $\Rightarrow$ **cross-validation**

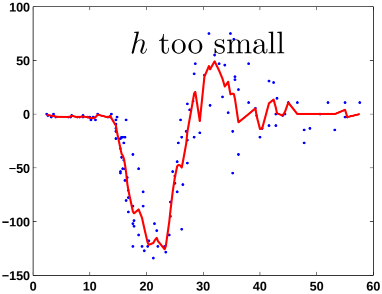
Effect of the tuning parameters

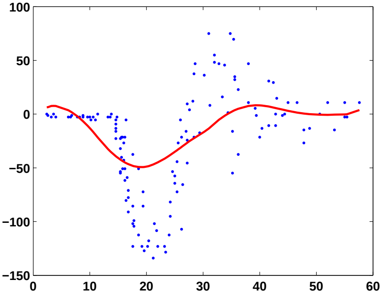


Effect of the tuning parameters

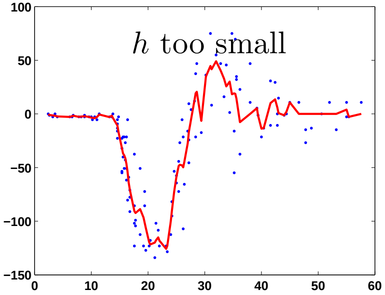


Effect of the tuning parameters

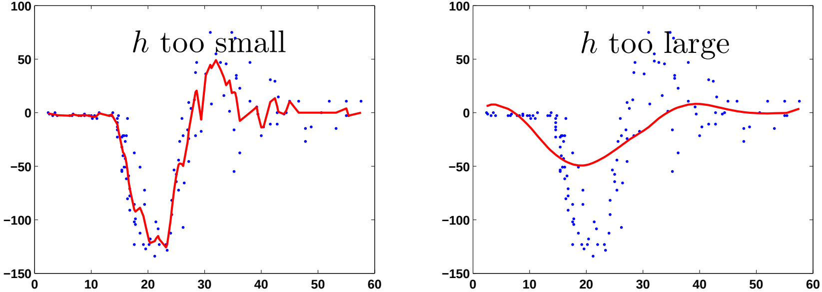




Effect of the tuning parameters



Effect of the tuning parameters



Model selection criteria are ABSOLUTELY needed

Outline

1. Goal & Overview
2. Introduction
   - Parametric vs. nonparametric regression
   - Nonparametric regression estimates: an overview
3. Fixed-Size Least Squares Support Vector Machines
   - Fixed Size LS-SVM formulation
   - Selection of Support Vectors
   - Practical identification problem
4. Robust Nonparametric Methods
   - Problems with outliers
   - Robust nonparametric regression
5. Correlated Errors
   - Problems with correlation in nonparametric regression
   - Removing correlation effects
6. Confidence Intervals
7. Conclusions

Estimation in Primal Space

- LS-SVM formulation for regression

Primal formulation

$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{k=1}^{n} e_k^2$$
$$s.t. \quad w^T\varphi(X_k) + b + e_k = Y_k, \quad k = 1,\ldots,n.$$

- Can we solve the LS-SVM in primal space instead of dual?
- Approximation of feature map φ needed
- Is it possible to compute such a mapping?
  - φ can be infinite dimensional
  - Solution:use a *fixed size* m of support vectors to approximate φ
  - Solve the above as primal ridge regression

Problems with Large Scale Data

1. Calculation and/or storage kernel matrix Ω
   - $N = 1.000 \Rightarrow \Omega \Rightarrow 8$ MB
   - $N = 10.000 \Rightarrow \Omega \Rightarrow 763$ MB
   - $N = 20.000 \Rightarrow \Omega \Rightarrow 3051$ MB
2. If possible to compute, how long would it take?

Problems with Large Scale Data

1. Calculation and/or storage kernel matrix Ω
   - $N = 1.000 \Rightarrow \Omega \Rightarrow 8$ MB
   - $N = 10.000 \Rightarrow \Omega \Rightarrow 763$ MB
   - $N = 20.000 \Rightarrow \Omega \Rightarrow 3051$ MB
2. If possible to compute, how long would it take?

$\Rightarrow$ Solution: Matrix Approximations (Nyström, 1930)

$$\hat{\varphi}_i(x) \overset{m \ll n}{\approx} \frac{\sqrt{m}}{\lambda_i^{(m)}} \sum_{k=1}^{m} K(X_k, x) u_{ki}^{(m)}$$

Fixed Size LS-SVM formulation

- Given: approximation to the feature map

Primal formulation

$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{k=1}^{n} e_k^2$$
$$s.t. \quad w^T \hat{\varphi}(X_k) + b + e_k = Y_k, \quad k = 1, \ldots, n.$$

- Solution

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

with

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

Selection of support vectors: Rényi Entropy

- Maximize quadratic Rényi entropy: $H_{R2}^{m} = -\log \int f(x)^2 \, dx$

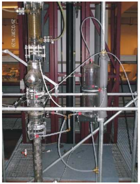
Theorem (Maximizing Entropy)

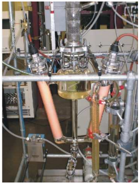
The Rényi entropy on a closed interval $[a, b]$ with $a, b \in \mathbb{R}$ and no additional moment constraints is maximized for the uniform density $1/(b-a)$.

(Selecting SV) (Wrong Bandwidth)

Identification of a pilot scale distillation column

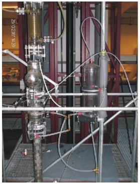
- Joint work with Bart Huyck (CIT)

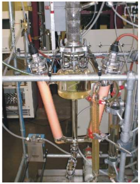




Identification of a pilot scale distillation column

- Joint work with Bart Huyck (CIT)

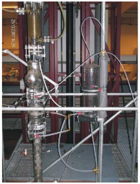


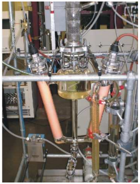


- Task: identify bottom temperature column with LS-SVM

Identification of a pilot scale distillation column

- Joint work with Bart Huyck (CIT)

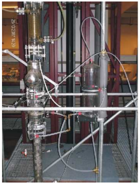


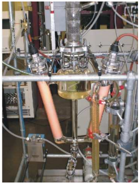


- Task: identify bottom temperature column with LS-SVM

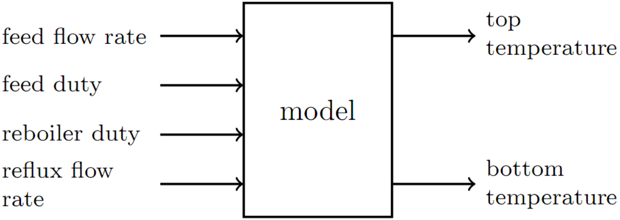
Identification of a pilot scale distillation column

- Joint work with Bart Huyck (CIT)



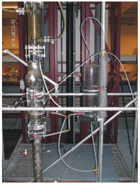


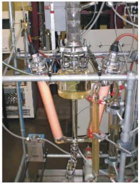
- Task: identify bottom temperature column with LS-SVM



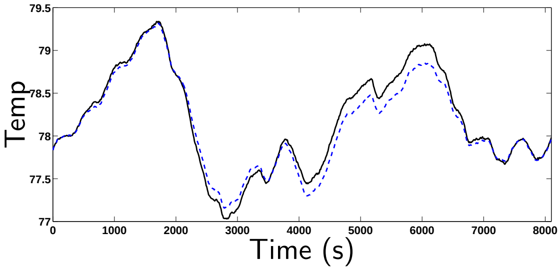
Identification of a pilot scale distillation column

- Joint work with Bart Huyck (CIT)





- Task: identify bottom temperature column with LS-SVM



Outline

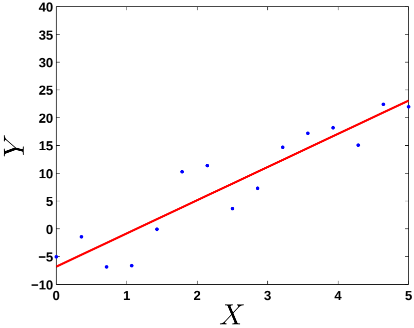
1. Goal & Overview
2. Introduction
   - Parametric vs. nonparametric regression
   - Nonparametric regression estimates: an overview
3. Fixed-Size Least Squares Support Vector Machines
   - Fixed Size LS-SVM formulation
   - Selection of Support Vectors
   - Practical identification problem
4. Robust Nonparametric Methods
   - Problems with outliers
   - Robust nonparametric regression
5. Correlated Errors
   - Problems with correlation in nonparametric regression
   - Removing correlation effects
6. Confidence Intervals
7. Conclusions

Problems with outliers in parametric regression

- model: $Y_k = aX_k + b + e_k, \quad k = 1, \dots, n$
- (a, b) estimated from data
- LS principle: $(\hat{a}, \hat{b}) = \underset{a,b \in \mathbb{R}^2}{\arg\min} \frac{1}{n} \sum_{k=1}^{n} [Y_k - (aX_k + b)]^2$

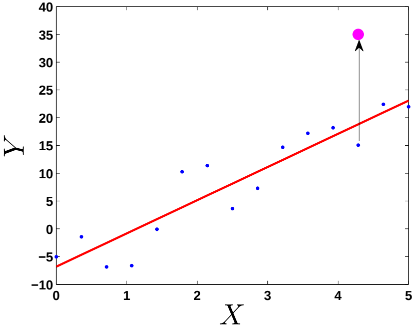
Problems with outliers in parametric regression

- model: $Y_k = aX_k + b + e_k, \quad k = 1, \ldots, n$
- (a, b) estimated from data
- LS principle: $(\hat{a}, \hat{b}) = \underset{a,b \in \mathbb{R}^2}{\arg\min} \frac{1}{n} \sum_{k=1}^{n} [Y_k - (aX_k + b)]^2$



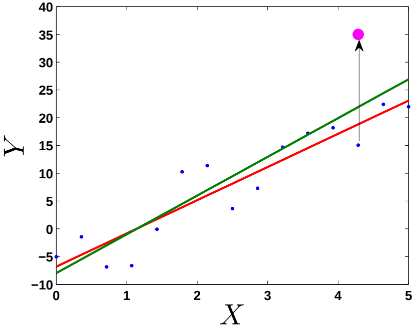
Problems with outliers in parametric regression

- model: $Y_k = aX_k + b + e_k, \quad k = 1, \ldots, n$
- (a, b) estimated from data
- LS principle: $(\hat{a}, \hat{b}) = \arg\min_{a,b \in \mathbb{R}^2} \frac{1}{n} \sum_{k=1}^{n} [Y_k - (aX_k + b)]^2$



Problems with outliers in parametric regression

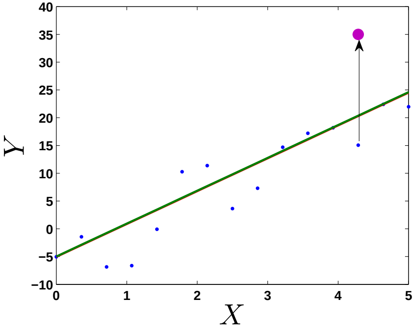
- model: $Y_k = aX_k + b + e_k, \quad k = 1, \ldots, n$
- (a, b) estimated from data
- LS principle: $(\hat{a}, \hat{b}) = \arg\min_{a,b \in \mathbb{R}^2} \frac{1}{n} \sum_{k=1}^{n} [Y_k - (aX_k + b)]^2$



- LS principle is NOT robust

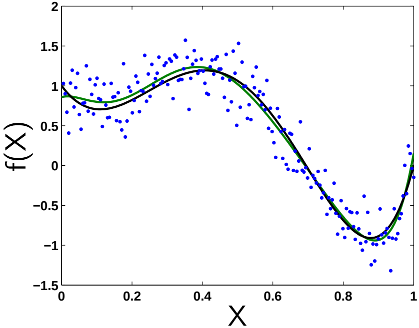
Problems with outliers in parametric regression

- model: $Y_k = aX_k + b + e_k, \quad k = 1, \ldots, n$
- (a, b) estimated from data
- LS principle: $(\hat{a}, \hat{b}) = \underset{a,b \in \mathbb{R}^2}{\arg\min} \frac{1}{n} \sum_{k=1}^{n} [Y_k - (aX_k + b)]^2$

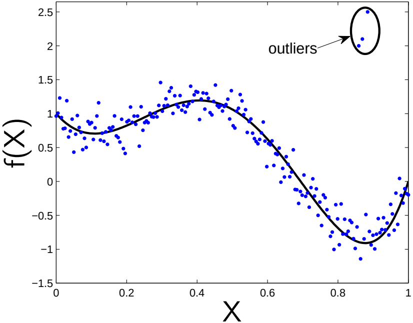


- LS principle is NOT robust
- Solution LAD: $(\hat{a}, \hat{b}) = \underset{a,b \in \mathbb{R}^2}{\arg\min} \frac{1}{n} \sum_{k=1}^{n} |Y_k - (aX_k + b)|$

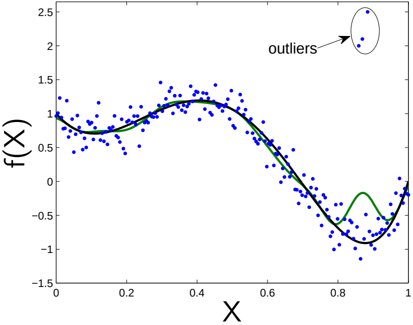
Problems with outliers in nonparametric regression



Problems with outliers in nonparametric regression



Problems with outliers in nonparametric regression



- LS principle ⇒ sensitive to outliers (and leverage points)
- Linear/polynomial kernel ⇒ non-robust methods
- Using appropriate CV ⇒ robust CV

Iterative Reweighting & Weight Functions

Primal formulation

$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{k=1}^{n} e_k^2$$
$$s.t. \quad Y_k = w^T \varphi(X_k) + b + e_k, \quad k = 1,\ldots,n.$$

Iterative Reweighting & Weight Functions

Primal formulation

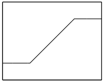
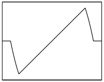
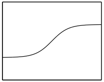
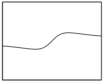
$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{k=1}^{n} v_k e_k^2$$
$$\text{s.t.} \quad Y_k = w^T \varphi(X_k) + b + e_k, \quad k = 1, \ldots, n.$$

Iterative Reweighting & Weight Functions

Primal formulation

$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{k=1}^{n} v_k e_k^2$$

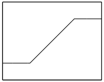
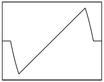
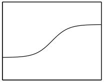
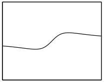
$$s.t. \quad Y_k = w^T \varphi(X_k) + b + e_k, \quad k = 1, \ldots, n.$$

	Huber	Hampel	Logistic	Myriad
$V(r)$	$\begin{cases} 1, & \text{if } \lvert r\rvert < \beta; \\ \frac{\beta}{\lvert r\rvert}, & \text{if } \lvert r\rvert \geq \beta. \end{cases}$	$\begin{cases} 1, & \text{if } \lvert r\rvert < b_1; \\ \frac{b_2 - \lvert r\rvert}{b_2 - b_1}, & \text{if } b_1 \leq \lvert r\rvert \leq b_2; \\ 0, & \text{if } \lvert r\rvert > b_2. \end{cases}$	$\frac{\tanh(r)}{r}$	$\frac{\delta^2}{\delta^2 + r^2}$
$\psi(r)$				
$L(r)$	$\begin{cases} r^2, & \text{if } \lvert r\rvert < \beta; \\ \beta\lvert r\rvert - \frac{1}{2}\beta^2, & \text{if } \lvert r\rvert \geq \beta. \end{cases}$	$\begin{cases} r^2, & \text{if } \lvert r\rvert < b_1; \\ \frac{b_2 r^2 - \lvert r^3\rvert}{b_2 - b_1}, & \text{if } b_1 \leq \lvert r\rvert \leq b_2; \\ 0, & \text{if } \lvert r\rvert > b_2. \end{cases}$	$r\tanh(r)$	$\log(\delta^2 + r^2)$

Iterative Reweighting & Weight Functions

Primal formulation

$$\min_{w,b,e} \mathcal{J}_P(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{k=1}^{n} {\color{red}v_k} e_k^2$$
$$s.t. \quad Y_k = w^T\varphi(X_k) + b + e_k, \quad k = 1,\ldots,n.$$

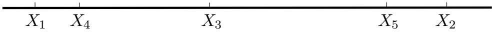
	Huber	Hampel	Logistic	Myriad
$V(r)$	$\begin{cases}1, & \text{if } \lvert r\rvert < \beta;\\ \frac{\beta}{\lvert r\rvert}, & \text{if } \lvert r\rvert \geq \beta.\end{cases}$	$\begin{cases}1, & \text{if } \lvert r\rvert < b_1;\\ \frac{b_2-\lvert r\rvert}{b_2-b_1}, & \text{if } b_1 \leq \lvert r\rvert \leq b_2;\\ 0, & \text{if } \lvert r\rvert > b_2.\end{cases}$	$\frac{\tanh(r)}{r}$	$\frac{\delta^2}{\delta^2+r^2}$
$\psi(r)$				
$L(r)$	$\begin{cases}r^2, & \text{if } \lvert r\rvert < \beta;\\ \beta\lvert r\rvert - \frac{1}{2}\beta^2, & \text{if } \lvert r\rvert \geq \beta.\end{cases}$	$\begin{cases}r^2, & \text{if } \lvert r\rvert < b_1;\\ \frac{b_2 r^2-\lvert r^3\rvert}{b_2-b_1}, & \text{if } b_1 \leq \lvert r\rvert \leq b_2;\\ 0, & \text{if } \lvert r\rvert > b_2.\end{cases}$	$r\tanh(r)$	$\log(\delta^2+r^2)$

Properties of the Myriad

- if $\delta \to \infty \Longrightarrow$ Myriad converges to sample mean
- if $\delta \to 0 \quad \Longrightarrow$ Myriad converges to the sample mode

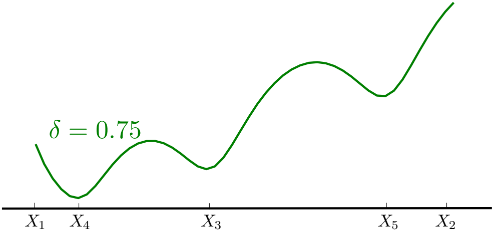
Properties of the Myriad

- if $\delta \to \infty \implies$ Myriad converges to sample mean
- if $\delta \to 0 \implies$ Myriad converges to the sample mode



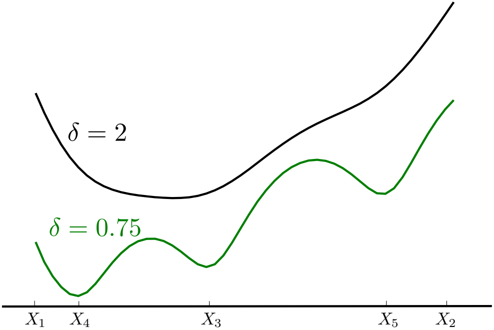
Properties of the Myriad

- if $\delta \to \infty \implies$ Myriad converges to sample mean
- if $\delta \to 0 \implies$ Myriad converges to the sample mode



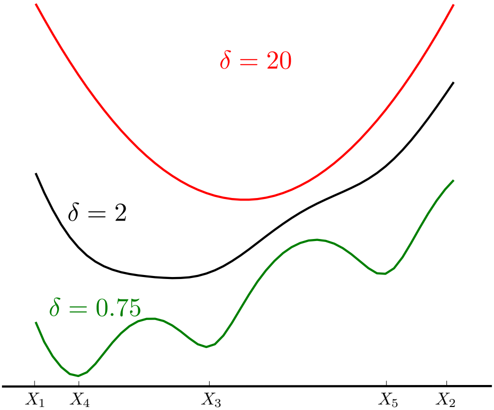
Properties of the Myriad

- if $\delta \to \infty \Longrightarrow$ Myriad converges to sample mean
- if $\delta \to 0 \Longrightarrow$ Myriad converges to the sample mode



Properties of the Myriad

- if $\delta \to \infty \implies$ Myriad converges to sample mean
- if $\delta \to 0 \implies$ Myriad converges to the sample mode



To obtain a fully robust solution...

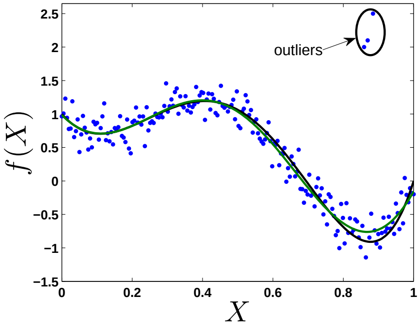
- robust smoother
- bounded kernel
- robust CV $\Rightarrow L'$ bounded

$$RCV(\theta) = \frac{1}{n} \sum_{i=1}^{n} L\left(Y_i - \hat{m}_n^{(-i)}(X_i;\theta)\right)$$

To obtain a fully robust solution...

- robust smoother
- bounded kernel
- robust CV $\Rightarrow L'$ bounded

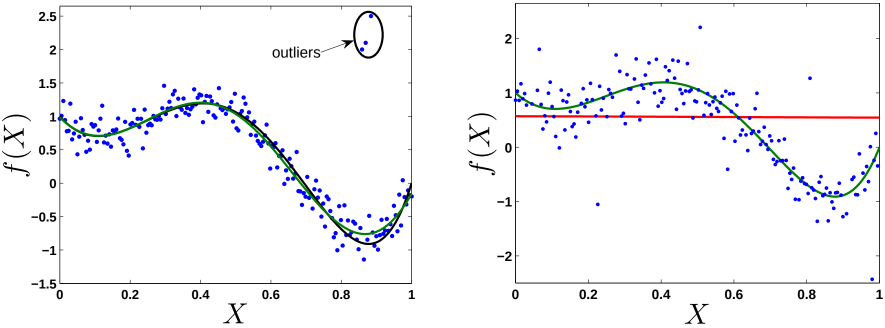
$$RCV(\theta) = \frac{1}{n}\sum_{i=1}^{n} L\left(Y_i - \hat{m}_n^{(-i)}(X_i;\theta)\right)$$



To obtain a fully robust solution...

- robust smoother
- bounded kernel
- robust CV $\Rightarrow L'$ bounded

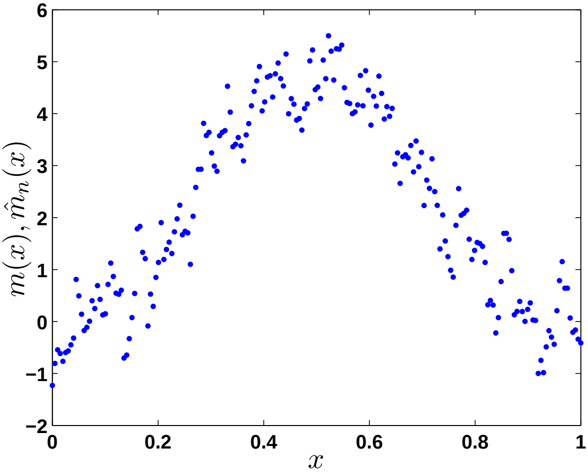
$$RCV(\theta) = \frac{1}{n}\sum_{i=1}^{n} L\left(Y_i - \hat{m}_n^{(-i)}(X_i;\theta)\right)$$



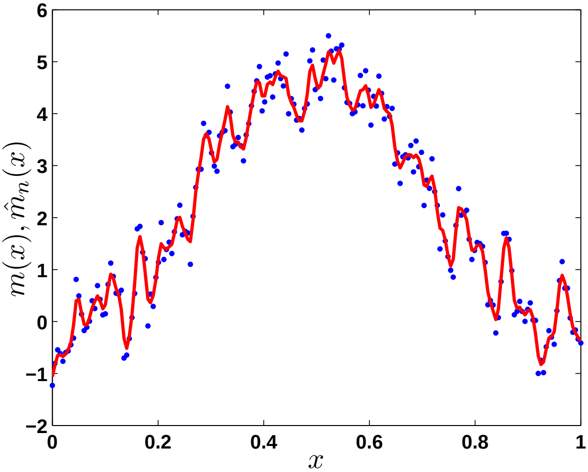
Outline

1. Goal & Overview
2. Introduction
   - Parametric vs. nonparametric regression
   - Nonparametric regression estimates: an overview
3. Fixed-Size Least Squares Support Vector Machines
   - Fixed Size LS-SVM formulation
   - Selection of Support Vectors
   - Practical identification problem
4. Robust Nonparametric Methods
   - Problems with outliers
   - Robust nonparametric regression
5. Correlated Errors
   - Problems with correlation in nonparametric regression
   - Removing correlation effects
6. Confidence Intervals
7. Conclusions

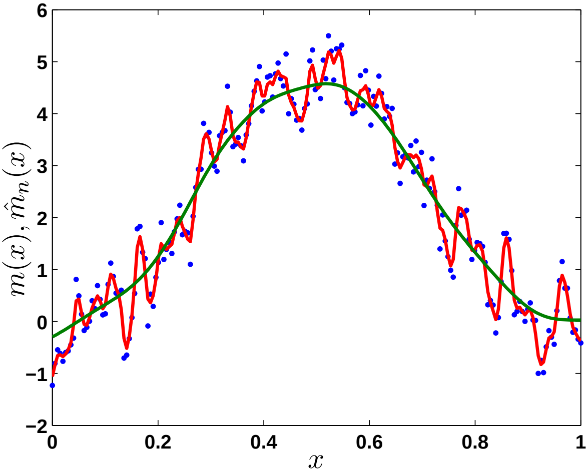
We have a problem!!!



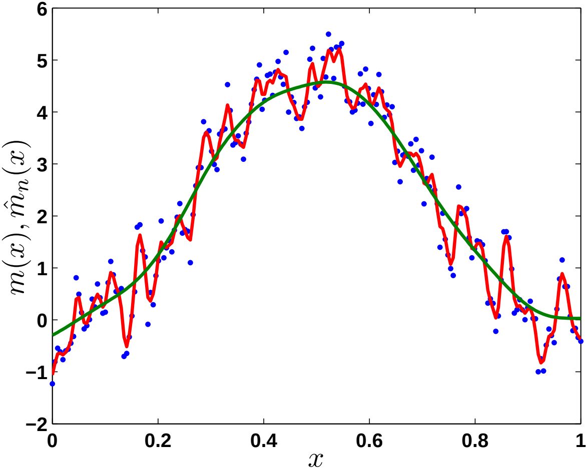
We have a problem!!!



We have a problem!!!



We have a problem!!!



VIOLATION OF I.I.D. ASSUMPTION

What went wrong?

- $Y_i = m(x_i) + e_i$: model selection $\Longrightarrow \text{Cov}[e_i, e_j] = 0$
- In previous example $\text{Cov}[e_i, e_j] \neq 0$
- correlation (covariance)? Strength of relationship between e_i and e_j

What went wrong?

- $Y_i = m(x_i) + e_i$: model selection $\Longrightarrow \mathrm{Cov}[e_i, e_j] = 0$
- In previous example $\mathrm{Cov}[e_i, e_j] \neq 0$
- correlation (covariance)? Strength of relationship between e_i and e_j

Example 1

Suppose there are two technology stocks. If they are affected by the same industry trends, their prices will tend to rise or fall together.

What went wrong?

- $Y_i = m(x_i) + e_i$: model selection $\Longrightarrow \text{Cov}[e_i, e_j] = 0$
- In previous example $\text{Cov}[e_i, e_j] \neq 0$
- correlation (covariance)? Strength of relationship between e_i and e_j

Example 1

Suppose there are two technology stocks. If they are affected by the same industry trends, their prices will tend to rise or fall together.

Example 2

Housing prices. Sensitive to offer and demand.

What went wrong?

- $Y_i = m(x_i) + e_i$: model selection $\Longrightarrow \mathrm{Cov}[e_i, e_j] = 0$
- In previous example $\mathrm{Cov}[e_i, e_j] \neq 0$
- correlation (covariance)? Strength of relationship between e_i and e_j

Example 1

Suppose there are two technology stocks. If they are affected by the same industry trends, their prices will tend to rise or fall together.

Example 2

Housing prices. Sensitive to offer and demand.

Example 3

In our toy example: e_{i+1} was affected by the value of e_i and so on...

Removing correlation effects: main theorem

Breakdown bandwidth selection procedures, smoother stays consistent!!

Theorem

Assume $x \equiv i/n$, $x \in [0,1]$, $\mathbf{E}[e] = 0$, $\text{Cov}[e_i, e_{i+k}] = \mathbf{E}[e_i e_{i+k}] = \gamma_k$ *and* $\gamma_k \sim k^{-a}$ *for some* $a > 2$. *Assume that* $Y_i = m(x_i) + e_i$ *and*

(C1) K *is Lipschitz continuous at* $x = 0$;

(C2) $\int K(u)\,du = 1$, $\lim_{|u|\to\infty} |uK(u)| = 0$, $\int |K(u)|\,du < \infty$, $\sup_u |K(u)| < \infty$;

(C3) $\int |k(u)|\,du < \infty$ *and* K *is symmetric.*

Further, assume that boundary effects are ignored and that $h \to 0$ *as* $n \to \infty$ *such that* $nh^2 \to \infty$, *then for the NW smoother it follows that*

$$\mathbf{E}[\text{CV}(h)] = \frac{1}{n}\mathbf{E}\sum_{i=1}^{n}\left[m(x_i) - \hat{m}_n^{(-i)}(x_i)\right]^2 + \sigma^2 - \frac{4K(0)}{nh - K(0)}\sum_{k=1}^{\infty}\gamma_k + o(n^{-1}h^{-1})$$

Removing correlation effects: main theorem

Breakdown bandwidth selection procedures, smoother stays consistent!!

Theorem

Assume $x \equiv i/n$, $x \in [0,1]$, $\mathbf{E}[e] = 0$, $\mathrm{Cov}[e_i, e_{i+k}] = \mathbf{E}[e_i e_{i+k}] = \gamma_k$ *and* $\gamma_k \sim k^{-a}$ *for some* $a > 2$. *Assume that* $Y_i = m(x_i) + e_i$ *and*

(C1) K *is Lipschitz continuous at* $x = 0$;

(C2) $\int K(u)\,du = 1$, $\lim_{|u|\to\infty} |uK(u)| = 0$, $\int |K(u)|\,du < \infty$, $\sup_u |K(u)| < \infty$;

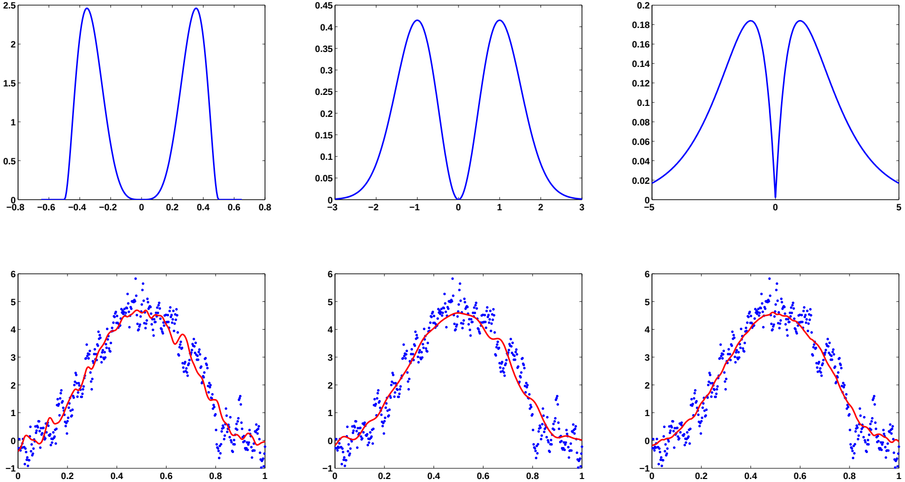
(C3) $\int |k(u)|\,du < \infty$ *and* K *is symmetric.*

Further, assume that boundary effects are ignored and that $h \to 0$ *as* $n \to \infty$ *such that* $nh^2 \to \infty$, *then for the NW smoother it follows that*

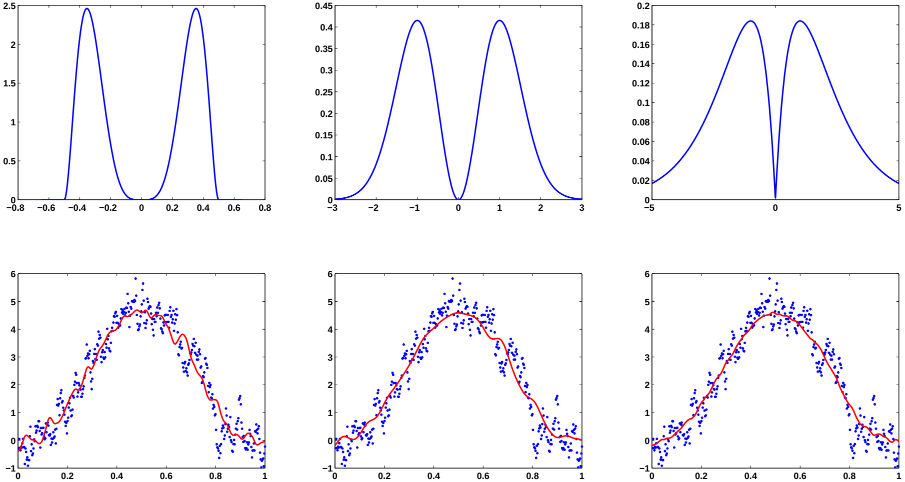
$$\mathbf{E}[\mathrm{CV}(h)] = \frac{1}{n}\mathbf{E}\sum_{i=1}^{n}\left[m(x_i) - \hat{m}_n^{(-i)}(x_i)\right]^2 + \sigma^2 - \frac{4K(0)}{nh - K(0)}\sum_{k=1}^{\infty}\gamma_k + o(n^{-1}h^{-1})$$

No prior knowledge about correlation structure needed !!

Suitable kernels & drawback



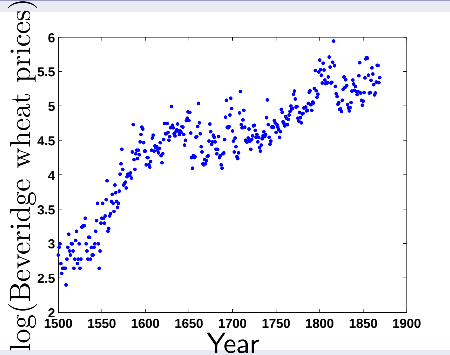
Suitable kernels & drawback



⇒⇒ Decreased Mean Squared Error ⇒⇒

Some real life examples

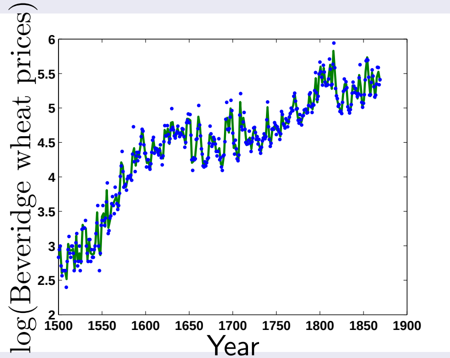
Beveridge index of wheat prices



U.S. monthly birth rate

Some real life examples

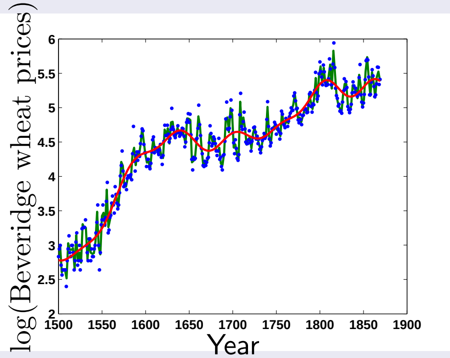
Beveridge index of wheat prices



U.S. monthly birth rate

Some real life examples

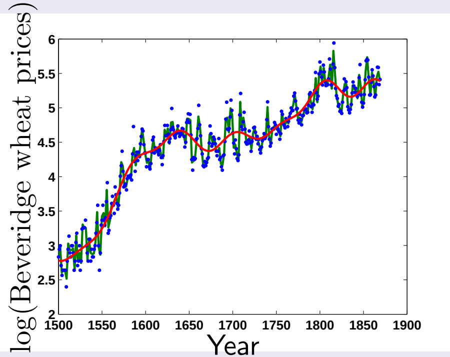
Beveridge index of wheat prices



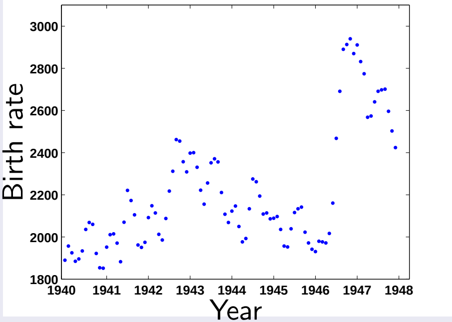
U.S. monthly birth rate

Some real life examples

Beveridge index of wheat prices

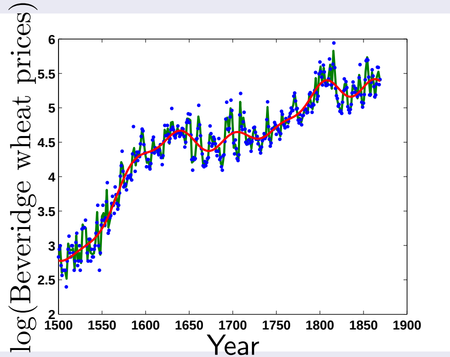


U.S. monthly birth rate

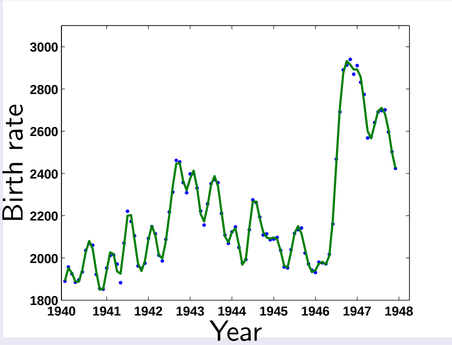


Some real life examples

Beveridge index of wheat prices

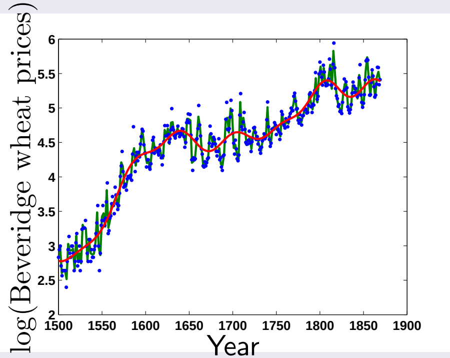


U.S. monthly birth rate

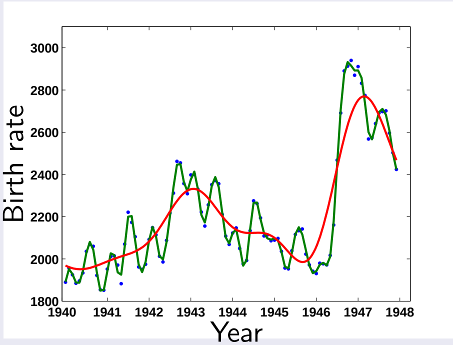


Some real life examples

Beveridge index of wheat prices



U.S. monthly birth rate



Outline

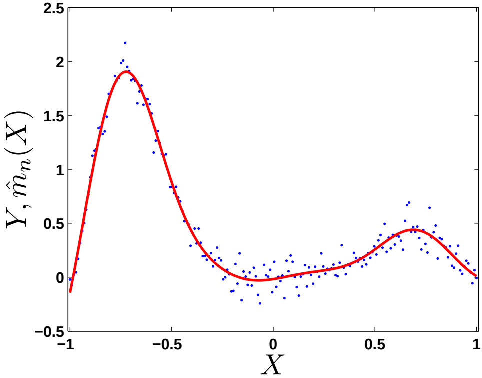
1. Goal & Overview
2. Introduction
   - Parametric vs. nonparametric regression
   - Nonparametric regression estimates: an overview
3. Fixed-Size Least Squares Support Vector Machines
   - Fixed Size LS-SVM formulation
   - Selection of Support Vectors
   - Practical identification problem
4. Robust Nonparametric Methods
   - Problems with outliers
   - Robust nonparametric regression
5. Correlated Errors
   - Problems with correlation in nonparametric regression
   - Removing correlation effects
6. Confidence Intervals
7. Conclusions

What are confidence intervals?

- How accurate are our nonparametric estimates?
- Can we say something about the true function m given $\hat{m}_n$?
- We want something of the form:
  $L_n(x) \leq m(x) \leq U_n(x) \quad \forall x$ for some confidence level α
- Pointwise vs. simultaneous/uniform confidence intervals

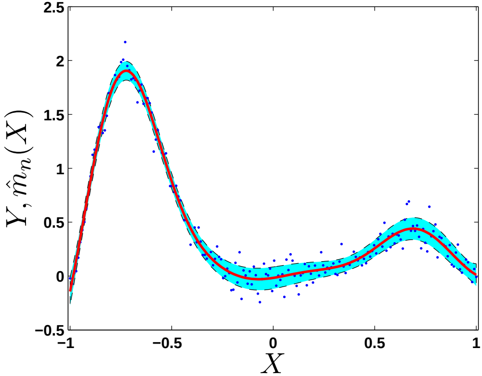
What are confidence intervals?

- How accurate are our nonparametric estimates?
- Can we say something about the true function m given $\hat{m}_n$?
- We want something of the form:
  $L_n(x) \leq m(x) \leq U_n(x) \quad \forall x$ for some confidence level α
- Pointwise vs. simultaneous/uniform confidence intervals



What are confidence intervals?

- How accurate are our nonparametric estimates?
- Can we say something about the true function m given $\hat{m}_n$?
- We want something of the form:
  $L_n(x) \leq m(x) \leq U_n(x) \quad \forall x$ for some confidence level α
- Pointwise vs. simultaneous/uniform confidence intervals



In practice...

In general

These intervals give the user the ability to see how well a certain model explains the true underlying process while taking statistical properties of the estimator into account.

Fault detection

In fault detection: CI are used for reducing the number of false alarms

Outline

1. Goal & Overview
2. Introduction
   - Parametric vs. nonparametric regression
   - Nonparametric regression estimates: an overview
3. Fixed-Size Least Squares Support Vector Machines
   - Fixed Size LS-SVM formulation
   - Selection of Support Vectors
   - Practical identification problem
4. Robust Nonparametric Methods
   - Problems with outliers
   - Robust nonparametric regression
5. Correlated Errors
   - Problems with correlation in nonparametric regression
   - Removing correlation effects
6. Confidence Intervals
7. Conclusions

Conclusions

Goal of the Thesis

Study the properties of LS-SVM for regression with an emphasis on statistical aspects and develop a framework for large scale data

Main Achievements

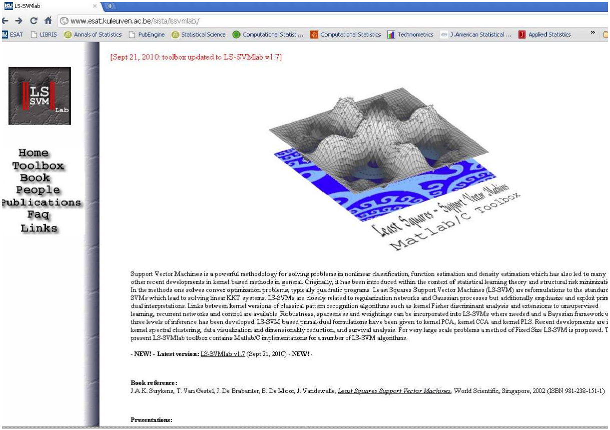
- Framework for large data sets
- Method for minimizing model selection criteria score functions
- Robustification of kernel based method
- Weight function with attractive properties
- Framework for correlated errors based on bimodal kernels
- Asymptotic normality of linear smoothers
- Bias & variance estimators for LS-SVM
- Pointwise & simultaneous CI + comparison bootstrap

LS-SVMLab software

- Free available (for research purposes) Matlab toolbox
- `http://www.esat.kuleuven.ac.be/sista/lssvmlab/`
- User's guide with applications (p. 113, De Brabanter *et al*, 2010)

LS-SVMLab software

- Free available (for research purposes) Matlab toolbox
- `http://www.esat.kuleuven.ac.be/sista/lssvmlab/`
- User's guide with applications (p. 113, De Brabanter *et al*, 2010)



Publications

- Falck, T., Dreesen, P., **De Brabanter, K.**, Pelckmans, K., De Moor, B., Suykens, J.A.K, Least-Squares Support Vector Machines for the Identification of Wiener-Hammerstein Systems, *Submitted*, 2011.
- **De Brabanter K.**, De Brabanter J., Suykens J.A.K., De Moor B., Kernel Regression in the Presence of Correlated Errors, *Submitted*, 2011.
- **De Brabanter K.**, Karsmakers P., De Brabanter J., Suykens J.A.K., De Moor B., Confidence Bands for Least Squares Support Vector Machine Classifiers: A Regression Approach, *Submitted*, 2010.
- **De Brabanter K.**, De Brabanter J., Suykens J.A.K., De Moor B., Approximate Confidence and Prediction Intervals for Least Squares Support Vector Regression, *IEEE Transactions on Neural Networks*, 22(1):110–120 , 2011.
- Sahhaf S., **De Brabanter K.**, Degraeve R., Suykens J.A.K., De Moor B., Groeseneken G., Modelling of Charge Trapping/De-trapping Induced Voltage Instability in High-k Gate Dielectrics, *Submitted*, 2010.
- Karsmakers P., Pelckmans K., **De Brabanter K.**, Van Hamme H., Suykens J.A.K., Sparse Conjugate Directions Pursuit with Application to Fixed-size Kernel Models, *Submitted*, 2010.
- Sahhaf S., Degraeve R., Cho M., **De Brabanter K.**, Roussel Ph.J., Zahid M.B., Groeseneken G., Detailed Analysis of Charge Pumping and $I_d - V_g$ Hysteresis for Profiling Traps in SiO_2/HfSiO(N), *Microelectronic Engineering*, 87(12):2614–2619, 2010.

Publications

- **De Brabanter K.**, De Brabanter J., Suykens J.A.K., De Moor B., Optimized Fixed-Size Kernel Models for Large Data Sets, *Computational Statistics & Data Analysis*, 54(6):1484–1504, 2010.
- López J., **De Brabanter K.**, Dorronsoro J.R., Suykens J.A.K., Sparse LS-SVMs with L_0-Norm Minimization, *Accepted for publication in Proc. of the 19th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Brugge (Belgium), 2011.
- Huyck B., **De Brabanter K.**, Logist F., De Brabanter J., Van Impe J., De Moor B., Identification of a Pilot Scale Distillation Column: A Kernel Based Approach, *Accepted for publication in 18th World Congress of the International Federation of Automatic Control (IFAC)*, 2011.
- **De Brabanter K.**, Karsmakers P., De Brabanter J., Pelckmans K., Suykens J.A.K., De Moor B., On Robustness in Kernel Based Regression, *NIPS 2010 Robust Statistical Learning (ROBUSTML) (NIPS 2010)*, Whistler, Canada, December 2010.
- **De Brabanter K.**, Sahhaf S., Karsmakers P., De Brabanter J., Suykens J.A.K., De Moor B., Nonparametric Comparison of Densities Based on Statistical Bootstrap, in *Proc. of the Fourth European Conference on the Use of Modern Information and Communication Technologies (ECUMICT)*, Gent, Belgium, March 2010, pp. 179–190.

Publications

- **De Brabanter K.**, De Brabanter J., Suykens J.A.K., De Moor B., Kernel Regression with Correlated Errors, in *Proc. of the the 11th International Symposium on Computer Applications in Biotechnology (CAB)*, Leuven, Belgium, July 2010, pp. 13–18.

- **De Brabanter K.**, Pelckmans K., De Brabanter J., Debruyne M., Suykens J.A.K., Hubert M., De Moor B., Robustness of Kernel Based Regression: a Comparison of Iterative Weighting Schemes, in *Proc. of the 19th International Conference on Artificial Neural Networks (ICANN)*, Limassol, Cyprus, September 2009, pp. 100–110.

- **De Brabanter K.**, Dreesen P., Karsmakers P., Pelckmans K., De Brabanter J., Suykens J.A.K., De Moor B., Fixed-Size LS-SVM Applied to the Wiener-Hammerstein Benchmark, in *Proc. of the 15th IFAC Symposium on System Identification (SYSID 2009)*, Saint-Malo, France, July 2009, pp. 826–831.

- **De Brabanter K.**, Karsmakers P., Ojeda F., Alzate C., De Brabanter J., Pelckmans K., De Moor B., Vandewalle J., Suykens J.A.K., LS-SVMlab Toolbox User's Guide version 1.7", Internal Report 10-146, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2010.