Sparse Regression Modelling Using an Incremental Weighted Optimization Method Based on Boosting with Correlation Criterion

S. Chen[†], X.X. Wang[‡] and D.J. Brown[‡]

[†] School of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.
E-mail: sqc@ecs.soton.ac.uk

[‡] Department of Creative Technologies
University of Portsmouth, Portsmouth PO1 3HE, U.K.
E-mails: xunxian.wang@port.ac.uk, david.j.brown@port.ac.uk

Presented at 10th Annual Conference of Chinese Automation and Computing Society in UK, Liverpool, UK, September 18, 2004

Motivation

Modelling from data: *generalization, interpretability, knowledge extraction* $\Rightarrow$ All depend on ability to construct appropriate sparse models

◯ Existing state-of-art sparse kernel regression modelling:

- Kernels position at training input data points with a common kernel variance

◯ This contribution considers generalized kernel model with tunable kernel centers and covariance matrices

$\Uparrow$ Enhancing modelling capability with much sparser representation

$\Downarrow$ Much more difficult nonlinear learning problem

- To manage learning complexity, incremental modelling is adopted to append kernel regressors one by one.

Generalized Kernel Modelling

- Modelling training data set $\{\mathbf{x}_l, y_l\}_{l=1}^N$ with regression model

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{M} w_i g_i(\mathbf{x})$$

- Generalized kernel

$$g_i(\mathbf{x}) = G\left(\sqrt{(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)}\right)$$

where $\boldsymbol{\mu}_i$ is kernel center and $\boldsymbol{\Sigma}_i$ diagonal kernel covariance matrix

- Define k-term model residuals over training set

$$y_i^{(k)} = y_i^{(k-1)} - w_k g_k(\mathbf{x}_i), \ 1 \le i \le N$$

Obviously $y_i^{(0)} = y_i$, the desired output

Incremental Modelling

- Mean square error of k-term regression model

$$MSE_k = \frac{1}{N} \sum_{i=1}^{N} \left(y_i^{(k-1)} - w_k g_k(\mathbf{x}_i) \right)^2$$

- k-th regression stage constructs the k-th regressor by determining:

kernel center $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$, as well as the usual LS weight solution

$$w_k = \frac{\sum_{i=1}^{N} y_i^{(k-1)} g_k(\mathbf{x}_i)}{\sum_{i=1}^{N} g_k^2(\mathbf{x}_i)}$$

- Model construction is terminated at M stage if

$$MSE_M < \xi$$

where ξ is a prescribed modelling accuracy, yielding an M-term generalized kernel model

Correlation Criterion

- Correlation between regressor $g_k(\mathbf{x})$ and training set $\{y_i^{(k-1)}, \mathbf{x}_i\}_{i=1}^N$

$$C_k(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{\sum_{i=1}^N g_k(\mathbf{x}_i) y_i^{(k-1)}}{\sqrt{\sum_{i=1}^N g_k^2(\mathbf{x}_i)} \sqrt{\sum_{i=1}^N \left(y_i^{(k-1)}\right)^2}}$$

defines similarity between regressor and training set

- Regressor positioning and shaping

$$\max_{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k} \left| C_k\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\right|$$

- It can be shown

$$\max_{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k} \left| C_k\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\right| \Leftrightarrow \min_{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k} MSE_k$$

Guided Random Search

Consider task of minimizing $f(\mathbf{u})$

Outer Loop: N_G number of generations

Initialization: keep best solution found in previous generation as $\mathbf{u}_1$ and randomly choose rest of population $\mathbf{u}_2, \cdots, \mathbf{u}_{P_S}$

Inner Loop: N_I iterations

- Perform a convex combination

$$\mathbf{u}_{P_S+1} = \sum_{i=1}^{P_S} \delta_i \mathbf{u}_i$$

- Weightings

$$\delta_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{P_S} \delta_i = 1$$

are adopted (boosting) to reflect goodness of $\mathbf{u}_i$

- $\mathbf{u}_{P_S+1}$ replaces worst member in population $\mathbf{u}_i$, $1 \leq i \leq P_S$

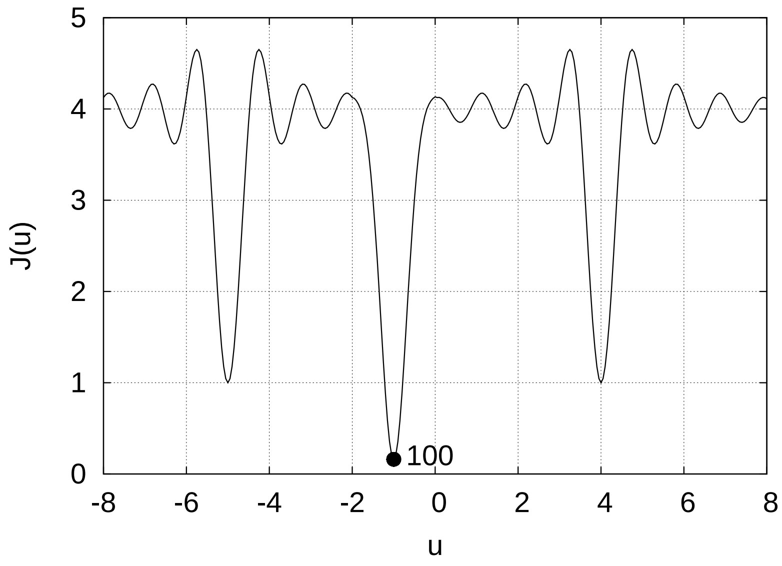
End of *Inner Loop*

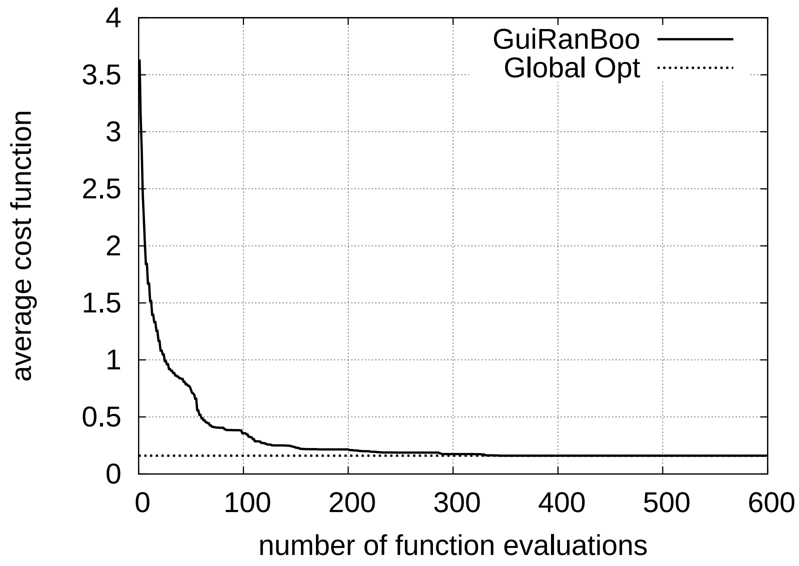
End of *Outer Loop*

Optimization Example

◯ Population size $P_S = 6$, number of Inner iterations $N_I = 20$ and number of generations $N_G = 12$

◯ 100 random experiments, populations of all 100 runs converge to global minimum





Simple Modelling Example

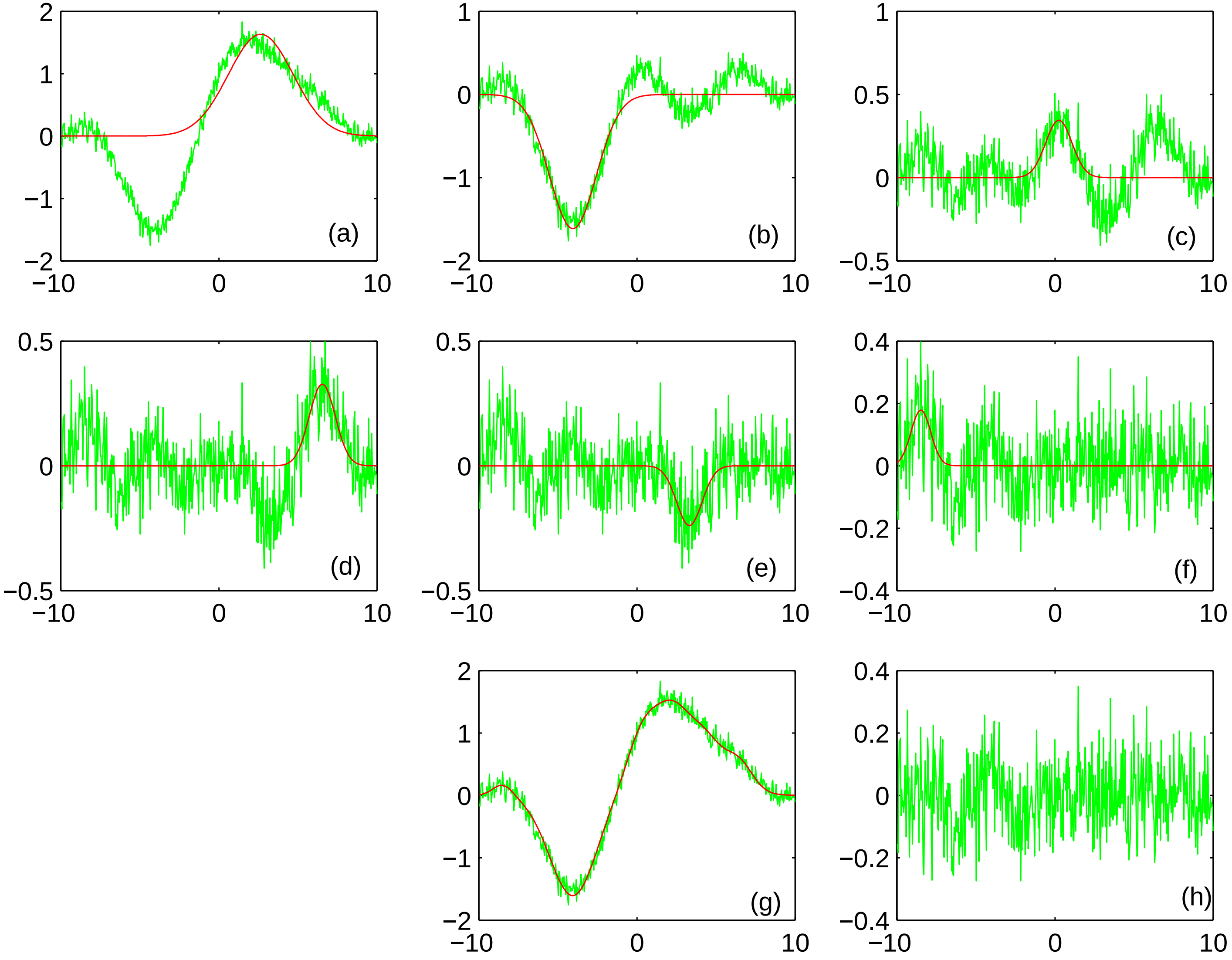
- 500 points of training data generated from

$$y(x) = 0.1x + \frac{\sin x}{x} + \sin 0.5x + \epsilon$$

where $x \in [-10,\ 10]$ and ϵ Gaussian white noise of variance 0.01

- Generalized Gaussian kernel used, modelling accuracy set to $\xi = 0.012$:

regression step k	mean μ_k	variance σ_k^2	weight w_k	MSE MSE_k
0	–	–	–	0.8431
1	2.6905	4.2488	1.6088	0.3703
2	-4.0837	2.1853	-1.6019	0.0341
3	0.2982	0.6000	0.3781	0.0243
4	6.6062	0.6610	0.3116	0.0173
5	3.4162	0.6091	-0.2242	0.0138
6	-8.4780	0.4295	0.1787	0.0119


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
−10
0
10

Gas Furnace Data Modelling

○ Modelling relationship between coded input gas feed rate (input $u(t)$) and CO_2 concentration from gas furnace (output $y(t)$):

Series J in: G.E.P. Box and G.M. Jenkins, *Time Series Analysis, Forecasting and Control*. Holden Day Inc., 1976.

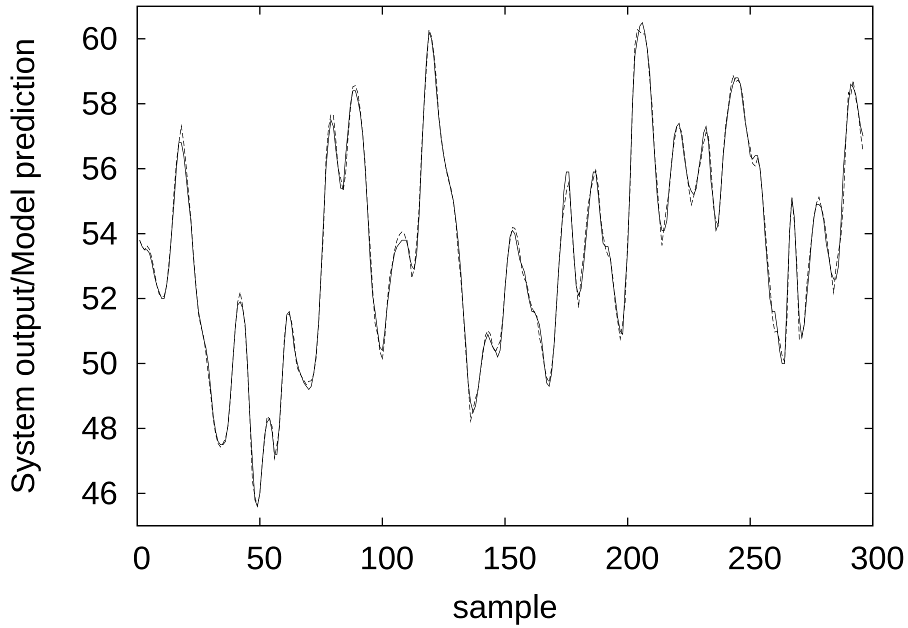
○ Data set contains 296 pairs of input-output samples (u_i, y_i), modelled as $y_i = f_s(\mathbf{x}_i) + \epsilon_i$ with

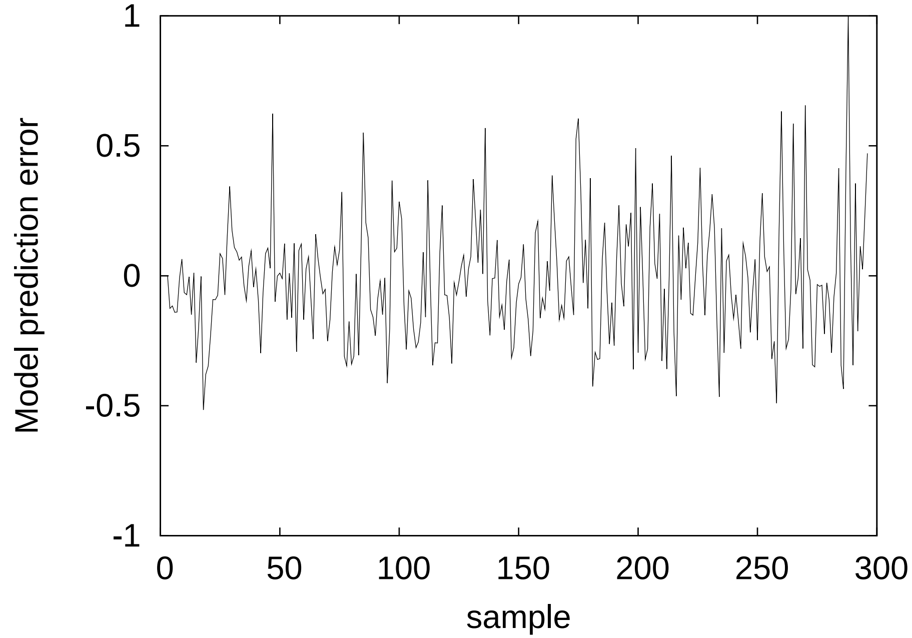
$$\mathbf{x}_i = [y_{i-1}\ y_{i-2}\ y_{i-3}\ u_{i-1}\ u_{i-2}\ u_{i-3}]^T$$

○ Generalized Gaussian kernel used, modelling accuracy set to $\xi = 0.054$: proposed incremental modelling method yields a 18-term generalized kernel model

○ To achieve same modelling accuracy for this data set, best of existing state-of-art kernel regression techniques required at least 28 regressors

Gas Furnace Data Modelling (continue)





Noisy training output data y_i, model output $\hat{y}_i$ and modelling error $\epsilon_i = y_i - \hat{y}_i$

Conclusions

- A novel construction algorithm has been proposed for parsimonious regression modelling based on generalized kernel model
- Proposed algorithm has ability to tune center and diagonal covariance matrix of individual regressor to incrementally maximize correlation criterion (minimize training mean square error)
- A guided random search method has been developed to append regressors one by one in an incremental modelling procedure
- Our method offers enhanced modelling capability with very sparse representation