Inductive Model Generation and Multimodel Selection

Vadim STRIZHOV

Computing Center of the Russian Academy of Sciences

14th April 2008, Paris

The primary goal of the research is

to develop the theory and the practice of the inductively-generated regression models

Applications:

Biology, Medicine, Ecology, Economics

Keywords: *mathematical modelling, regression analysis*

Context and state of the art

The following well-developed techniques are involved:

1. Group Method of Data Handling
   [Ivakhnenko, A. G., Malada, H. R.]
2. Genetic programming
   [Koza, J. R., Zelinka, I.]
3. Optimal Brain Surgery
   [LeCun, Y., Solla, S. A.]
4. Model Selection and Coherent Bayesian Inference
   [Bishop, C., Nabney, J.]
5. Minimum Description Length Principle
   [MacKay, D., Grunwald, P D.]

Let there be given

The sample set:

$\{\mathbf{x}_1, ..., \mathbf{x}_N | \mathbf{x} \in \mathbb{R}^P\}$ the independent variables

$\{y_1, ..., y_N | y \in \mathbb{R}\}$ the corresponding depended variables

denote by D the data set $\{(\mathbf{x}_i, y_i)\}$.

The primitive functions:

$G = \{g | g : \mathbb{R} \times ... \times \mathbb{R} \longrightarrow \mathbb{R}\}$ smooth parametric functions

$g = g(\mathbf{b}, \cdot, \cdot, ..., \cdot)$

G defines the set of arbitrary superpositions $\mathcal{F} = \{f_i\}$
inductively by its elements g

$f_i = f_i(\mathbf{w}, \mathbf{x})$

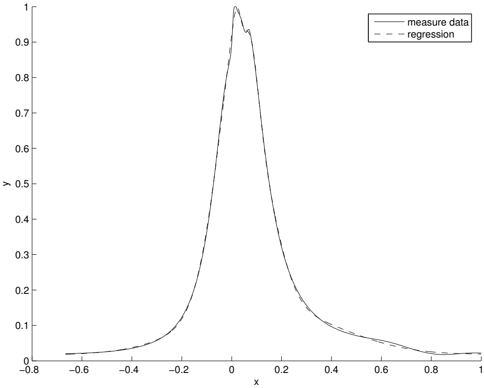
where $\mathbf{w} = \mathbf{b}_1 \vdots \mathbf{b}_2 \vdots ... \vdots \mathbf{b}_r$

Model of the optimal structure to be found

$$y = f_i(\mathbf{w}, \mathbf{x}) + \nu$$

One must find the model f_i, which brings the maximum to the target function $P(\mathbf{w}|D, \alpha, \beta, f_i)$

Practical example



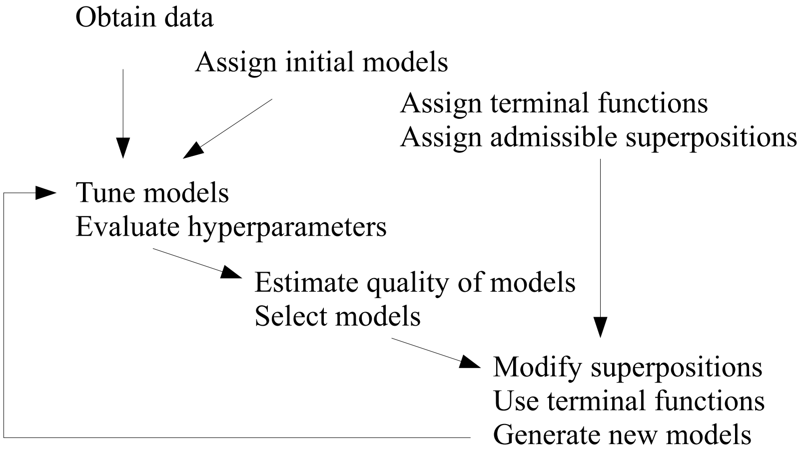
The pressure in the combusting camera of the diesel engine

x — crankshaft rotation angle, normalized
y — pressure, normalized
the data set contain 4000 samples

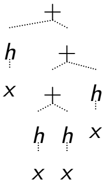
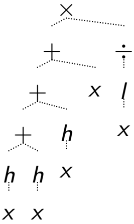
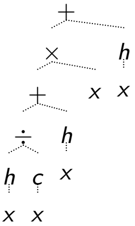
Primitive functions

Function	Description	Parameters
	$g(\mathbf{b}, x_1, x_2)$	
plus	$y = x_1 + x_2$	–
times	$y = x_1 x_2$	–
	$g(\mathbf{b}, x_1)$	
divide	$y = 1/x$	–
multiply	$y = ax$	a
add	$y = x + a$	a
gaussian	$y = \frac{\lambda}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\xi)^2}{2\sigma^2}\right) + a$	λ, σ, ξ, a
linear	$y = ax + b$	a, b
parabolic	$y = ax^2 + bx + c$	a, b, c
cubic	$y = ax^3 + bx^2 + cx + d$	a, b, c, d
logsig	$y = \frac{\lambda}{1+\exp(-\sigma(x-\xi))} + a$	λ, σ, ξ, a

The process of the model construction



Selected models

Model 1	Model 2	Model 3
		

Legend: h — gaussian $y = \lambda(2\pi\sigma^{-1/2})\exp(-(x-\xi)^2(2\sigma^{-2}) + a)$,
c — cubic $y = ax^3 + bx^2 + cx + d$, l — linear $y = ax + b$.

$$f_2 = g_1(g_2(g_3(g_4(g_5(x), g_6(x)), g_7(x)), x), g_8(x)).$$

The full representation of the Model 2

$$y = (ax+b)^{-1}\left(x + \sum_{i=1}^{3} \frac{\lambda_i}{\sqrt{2\pi\sigma_i}}\exp\left(-\frac{(x-\xi_i)^2}{2\sigma_i^2}\right) + a_i\right).$$

Approach to the model selection

$f_1, \ldots, f_M$ are the competitive models,
$P(f_i|D)$ is the posterior probability, $P(D|f_i)$ is the evidence

$$P(f_i|D) = \frac{P(f_i)P(D|f_i)}{\sum_{j=1}^{M} P(D|f_j)P(f_j)}. \tag{1}$$

The models f_i and f_j could be compared as

$$\frac{P(f_i|D)}{P(f_j|D)} = \frac{P(f_i)P(D|f_i)}{P(f_j)P(D|f_j)}.$$

The posterior probability of the parameters $\mathbf{w}$ given D

$$P(\mathbf{w}|D, f_i) = \frac{P(D|\mathbf{w}, f_i)P(\mathbf{w}|f_i)}{P(D|f_i)}, \tag{2}$$

the model evidence in the parameter space is

$$P(D|f_i) = \int P(D|\mathbf{w}, f_i)P(\mathbf{w}|f_i)d\mathbf{w}.$$

Data generation hypothesis

$$y = f_i(\mathbf{w}, \mathbf{x}) + \nu,$$

the likelihood function is

$$P(y|\mathbf{x}, \mathbf{w}, \beta, f_i) \triangleq P(D|\mathbf{w}, \beta, f) = \exp(-\beta E_D(D|\mathbf{w}, f_i)) Z_D^{-1}(\beta),$$

the regularization function

$$P(\mathbf{w}|\alpha, f_i) = \exp(-\alpha E_W(\mathbf{w}|f_i)) Z_W^{-1}(\alpha),$$

$\beta = \sigma_\nu^{-2}$ the variance of data noise, $\alpha = \sigma_{\mathbf{w}}^{-2}$ the variance of parameters.

The desired target function

$$P(\mathbf{w}|D, \alpha, \beta, f_i) = \frac{P(D|\mathbf{w}, \beta) P(\mathbf{w}|\alpha)}{P(D|\alpha, \beta)} = \frac{\exp(-S(\mathbf{w}|f_i))}{Z_S(\alpha, \beta)}$$

and the error function $S(\mathbf{w}) = \alpha E_W + \beta E_D$.

Theoretical results

- The inductive model generation and multimodel selection method was introduced.
- The algorithm of the model modification using hyperparameters was proposed.
- Expert estimations concordance method for regression analysis was developed.

The software for generation and selection the nonlinear parametric models of the optimal complexity was developed.

Approvement of the project

- The project was supported by Russian Foundation of Basic Research, 2004–2006, 2005–2006, 2007–2008.
- The results were reported at the conferences:
  - Mathematical Methods of Pattern Recognition 2003, 2005, 2007;
  - Intellectual Information Analysis 2002, 2006;
  - Mathematics. Computer. Education. 2004, 2006, 2008.
- 12 peer-reviewed papers devoted to the subject were published.
- The lecture course created and delivered since 2006.
- The group of students are working on the theory and practice of the subject.

Practical applications

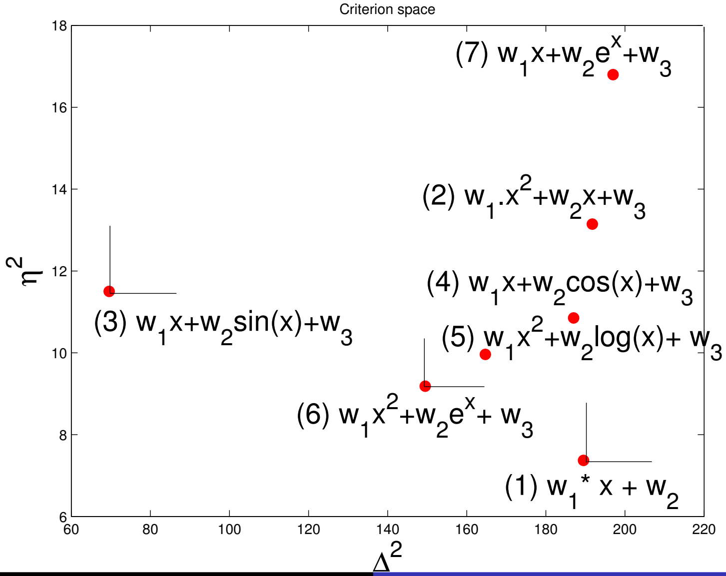
- ✓ Inductive generated models for regression analysis (Computing Center of RAS)
- ✓ The model of the pressure in the combusting camera (STMicroelectronics)
  the model of the oxygen sensor at the exhaust manifold of the diesel engine
- ✓ The volatility smile model of the option price at the stock markets (Forecsys Ltd.)
- ✓ Revelation behavior patterns of institutional owners at the world markets (RusAtom)
- ✓ The decision support system for the Russian Electricity Generation Industry (RAO ES)
- ✓ The mathematical model of the biomarkers of patients with CVD (ImmunoClin SARL)

Data generation hypothesis

The connections between data and parameter space are analyzed using Two-level Bayesian Inference.

- **The problem of dependencies between target functions and probability distribution functions for the data generation hypothesis is new.**

A model is the vector in the space of its criterions

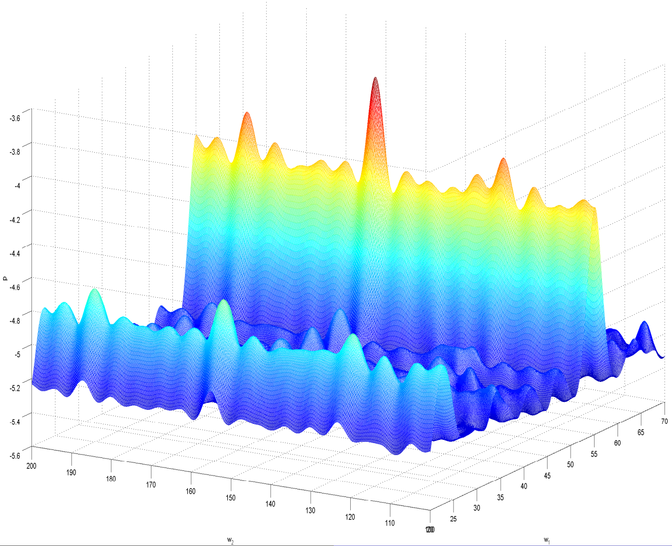


Parameter space analysis

The target functions define strategies of model parameter optimization and in the long run the most adequate model selection.

- **The optimal superposition structure could be discovered using the parameter space.**

Model stability and fitness in the parameter space

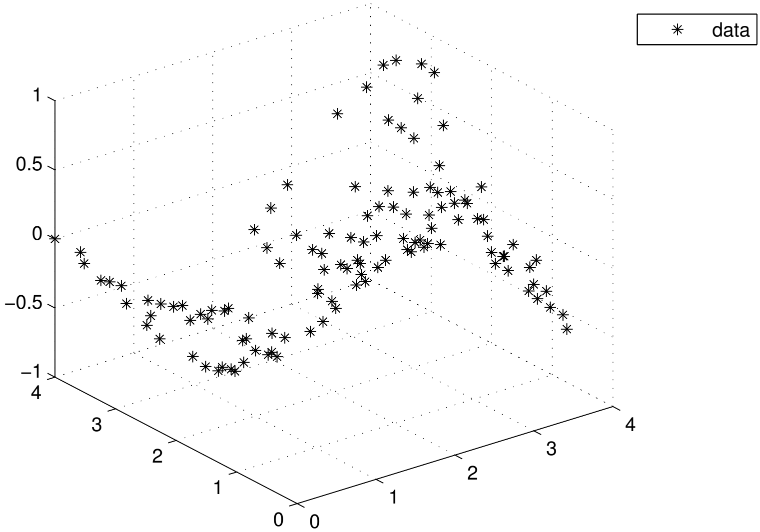


Direct model search and superposition structures

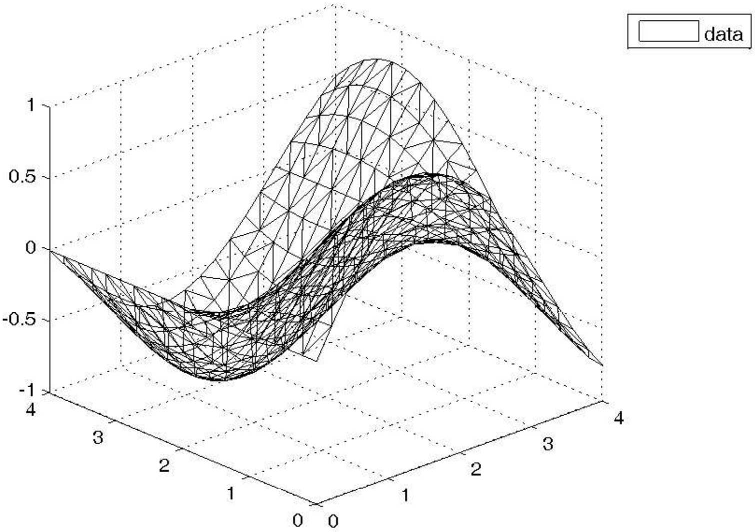
Now the stochastic optimization algorithm makes the model of the optimal structure.

- **There is a way to make the direct search of the optimal model.**

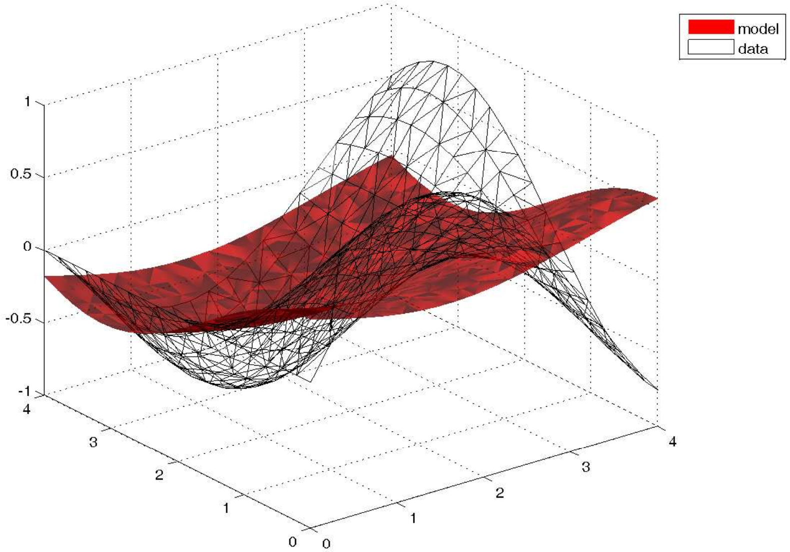
Direct search of the optimal superposition



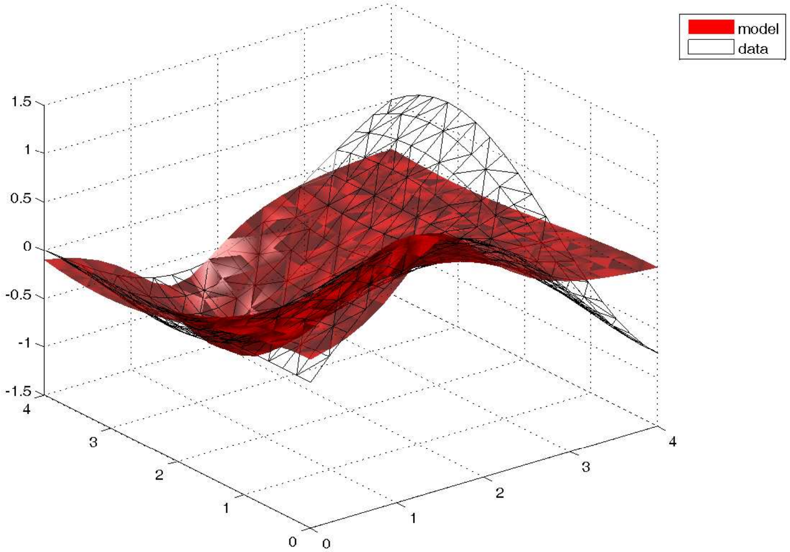
Direct search of the optimal superposition



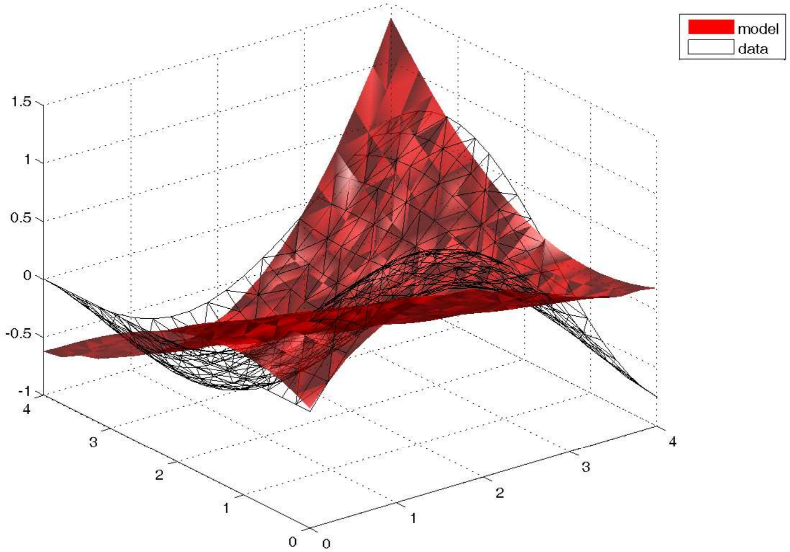
Direct search of the optimal superposition



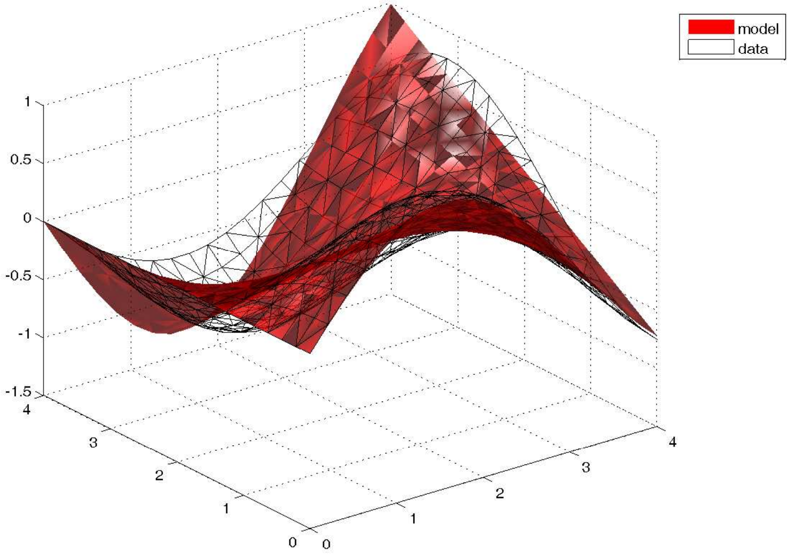
Direct search of the optimal superposition



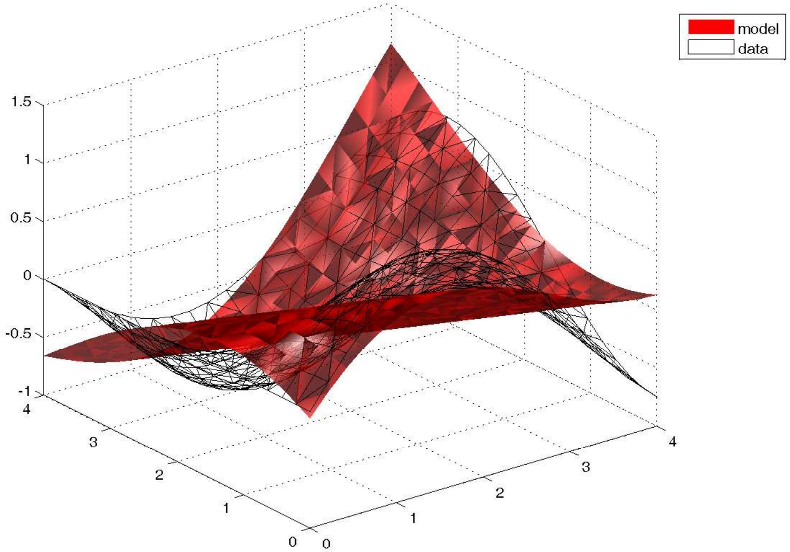
Direct search of the optimal superposition



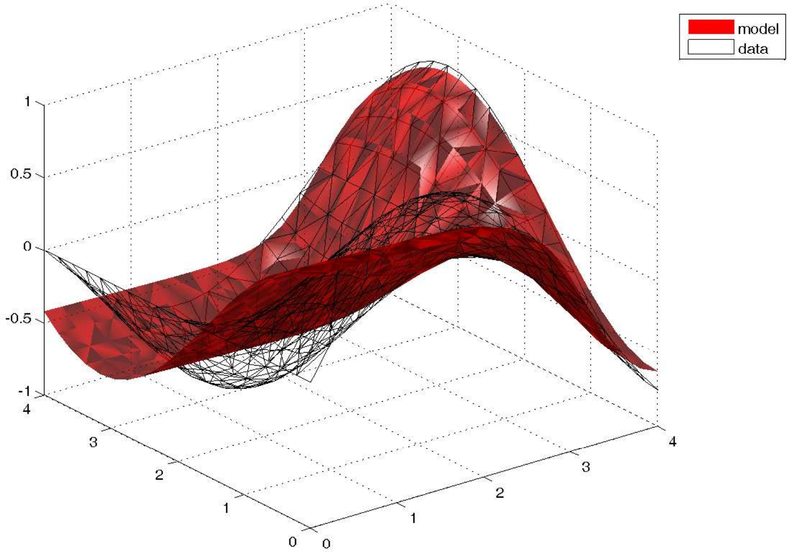
Direct search of the optimal superposition



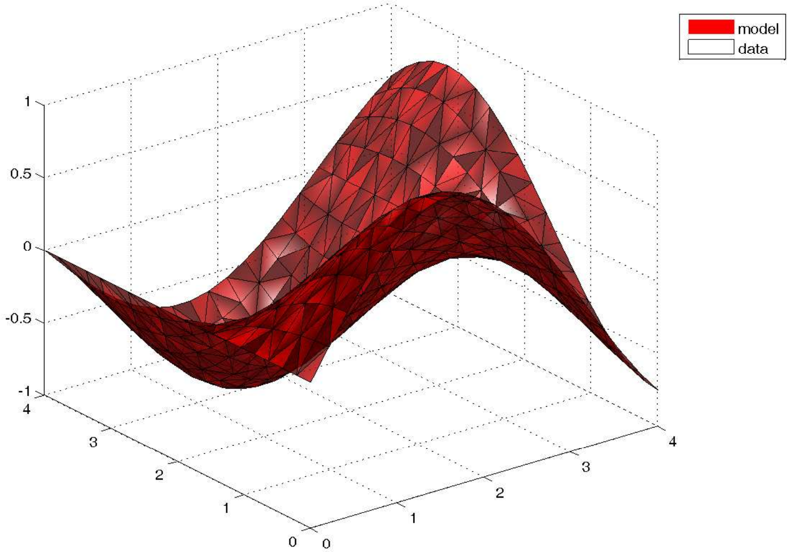
Direct search of the optimal superposition



Direct search of the optimal superposition



Direct search of the optimal superposition



Conclusion: Inductive Model Generation and Multimodel Selection

- The primary goal of this project is to develop the theory and the practice of the inductively-generated regression models.
- The subject is actual and asked-for. The project is approved by the experts.
- The project creates connections between researchers, who develop mathematical models for specific purposes.