Поиск оптимальной модели при решении задач нелинейной регрессии

Стрижов В.В.
Вычислительный центр им. А.А. Дородницына РАН

Дано

Задана выборка D:

$\{\mathbf{x}_1, ..., \mathbf{x}_N | \mathbf{x} \in \mathbb{R}^M\}$ — значения свободных переменных,
$\{y_1, ..., y_N | y \in \mathbb{R}\}$ — значения зависимой переменной.

Задано множество параметрических гладких функций

$G = \{g | g : \mathbb{R} \times ... \times \mathbb{R} \longrightarrow \mathbb{R}\}$,
$g = g(\mathbf{b}, \cdot, \cdot, ..., \cdot)$.

G задает множество суперпозиций $\mathcal{F} = \{f(\mathbf{w}, \mathbf{x})\}$ путем обобщенного индуктивного определения. Вектор параметров $\mathbf{w} = \mathbf{b}_1 \vdots \mathbf{b}_2 \vdots ... \vdots \mathbf{b}_r, \mathbf{w} \in \mathbb{R}^W$.

Найти

Регрессионную модель $f_i \in \mathcal{F}$, которая доставляет максимум

$$p(\mathbf{w}|D, \alpha, \beta, f_i).$$

α и β — дополнительные параметры.

Модель порождения данных

$y = f_i(\mathbf{w}, \mathbf{x}) + \nu$ — модель с аддитивным Гауссовским шумом с дисперсией σ_ν, $\beta = \frac{1}{\sigma_\nu^2}$.

Плотность вероятности появления данных

$$p(D|\mathbf{w}, \beta, f_i) = \frac{\exp(-\beta E_D(D|\mathbf{w}, f_i))}{Z_D(\beta)},$$

где

$$E_D = \frac{\beta}{2} \sum_{n=1}^{N} (f_i(\mathbf{x}_n) - y_n)^2.$$

Модель параметров

α — регуляризующий параметр, $\alpha = \frac{1}{\sigma_{\mathbf{w}}^2}$.

Плотность вероятности параметров

$$p(\mathbf{w}|\alpha, f_i) = \frac{\exp(-\alpha E_W(\mathbf{w}|f_i))}{Z_W(\alpha)},$$

где

$$E_W = \frac{\alpha}{2}||\mathbf{w}||^2.$$

Критерий поиска

При заданных значениях гиперпараметров α, β и фиксированной функции f_i

$$p(\mathbf{w}|D, \alpha, \beta, f_i) = \frac{p(D|\mathbf{w}, \beta, f_i)p(\mathbf{w}|\alpha, f_i)}{p(D|\alpha, \beta)} =$$

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

$S(\mathbf{w}|f_i) = \alpha E_W + \beta E_D$ — целевая функция.

NB: Связный Байесовский вывод

Первый уровень

$$p(\mathbf{w}|D, f_i) = \frac{p(D|\mathbf{w}, f_i)p(\mathbf{w}|f_i)}{\int p(D|\mathbf{w}', f_i)p(\mathbf{w}'|f_i)d\mathbf{w}'}$$

Второй уровень

$$P(f_i|D) = \frac{p(D|f_i)P(f_i)}{\sum_{i=1}^{M} p(D|f_i)P(f_i)}$$

Аппроксимация ошибки

Приблизим $S(\mathbf{w})$ рядом Тейлора второго порядка

$$S(\mathbf{w}) \approx S(\mathbf{w}^{\mathsf{MP}}) + \frac{1}{2}(\mathbf{w} - \mathbf{w}^{\mathsf{MP}})^T A(\mathbf{w} - \mathbf{w}^{\mathsf{MP}}),$$

матрица Гессе функции ошибок

$$A = \nabla^2 S(\mathbf{w}^{\mathsf{MP}}) = \beta\nabla^2 E_D(\mathbf{w}^{\mathsf{MP}}) + \alpha I.$$

Настройка гиперпараметров

$$2\alpha E_W^{\mathsf{MP}} = \gamma, \qquad 2\beta E_D^{\mathsf{MP}} = N - \gamma,$$

$$\gamma = W - \sum_{i=1}^{W} \frac{\lambda_i}{\lambda_i + \alpha},$$

$\lambda_1, ..., \lambda_W$ — собственные значения $\beta \nabla^2 E_D(\mathbf{w}^{\mathsf{MP}})$.

Алгоритм поиска

Заданы D, G и порожден набор $F_0 = \{f_1, ..., f_M\}$

Начало итераций

Для каждой f_i минимизируем функцию ошибок $S_i(\mathbf{w})$.

Определяем новые значения гиперпараметров α'_{ij} и β'_i

$$\beta'_i = \frac{N-\gamma}{E_D(f_i)}, \qquad \alpha'_{ij} = \frac{\gamma}{E_W(\mathbf{b}_{ij})}.$$

Изменение гиперпараметров может быть повторено после нахождения нового локального минимума $S_i(\mathbf{w})$.

Порождение моделей $f'_1, ..., f'_M$

Для каждой модели f_i строится производная модель f'_i.

В f_i выбирается элемент $g_{ij} : \min\limits_j(\alpha_{ij})$.

Выбирается модель f_ξ из $F_0 \setminus \{f_i\}$ и ее элемент $g_{\xi\zeta}$. Законы распределения величин ξ, ζ являются параметрами алгоритма оптимизации.

Модель f'_i порождается из модели f_i путем замещения элемента g_{ij} с его аргументами на элемент $g_{\xi\zeta}$ с его аргументами.

Мутация и селекция моделей

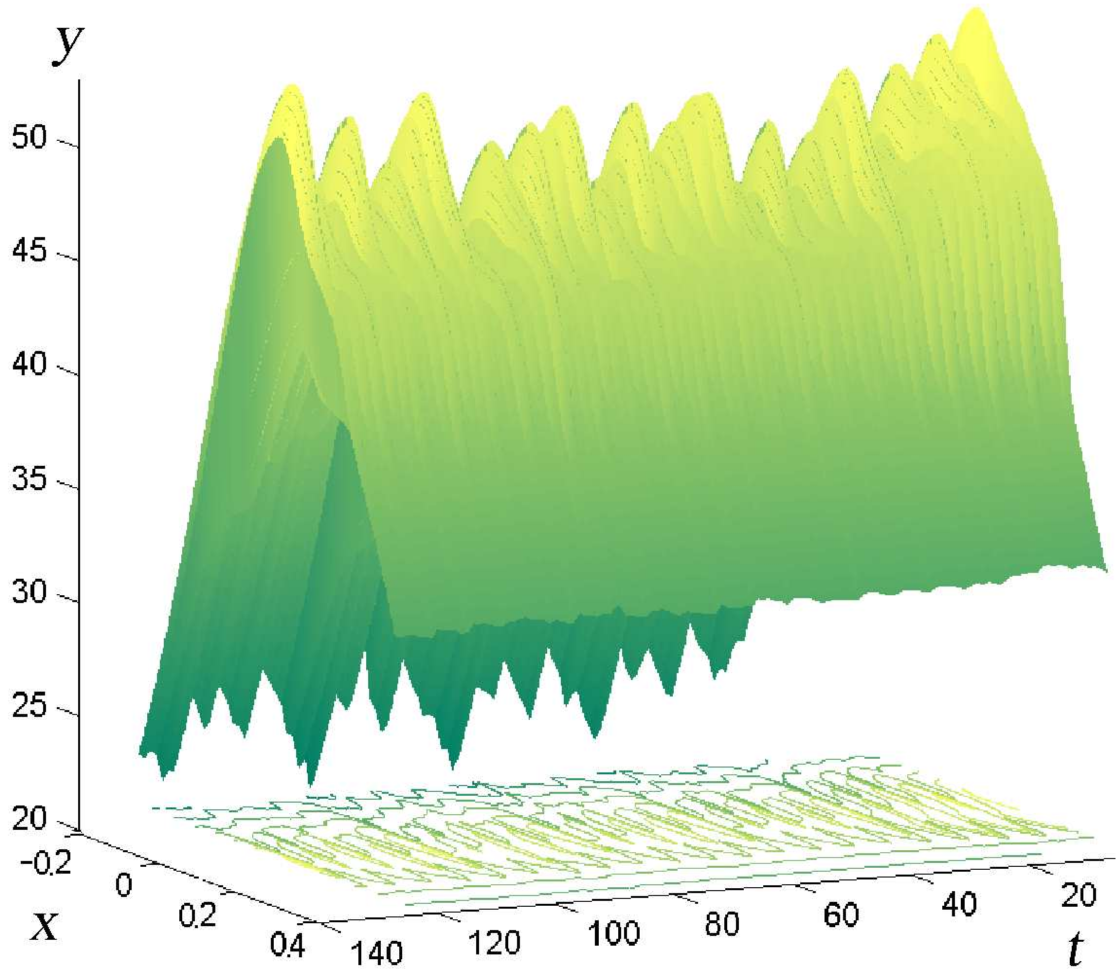
С заданной вероятностью f'_i подвергается мутации.

Из множества G выбирается элемент g_χ и замещает элемент $g'_{i\psi}$. Законы распределения величин χ, ψ — параметры алгоритма.

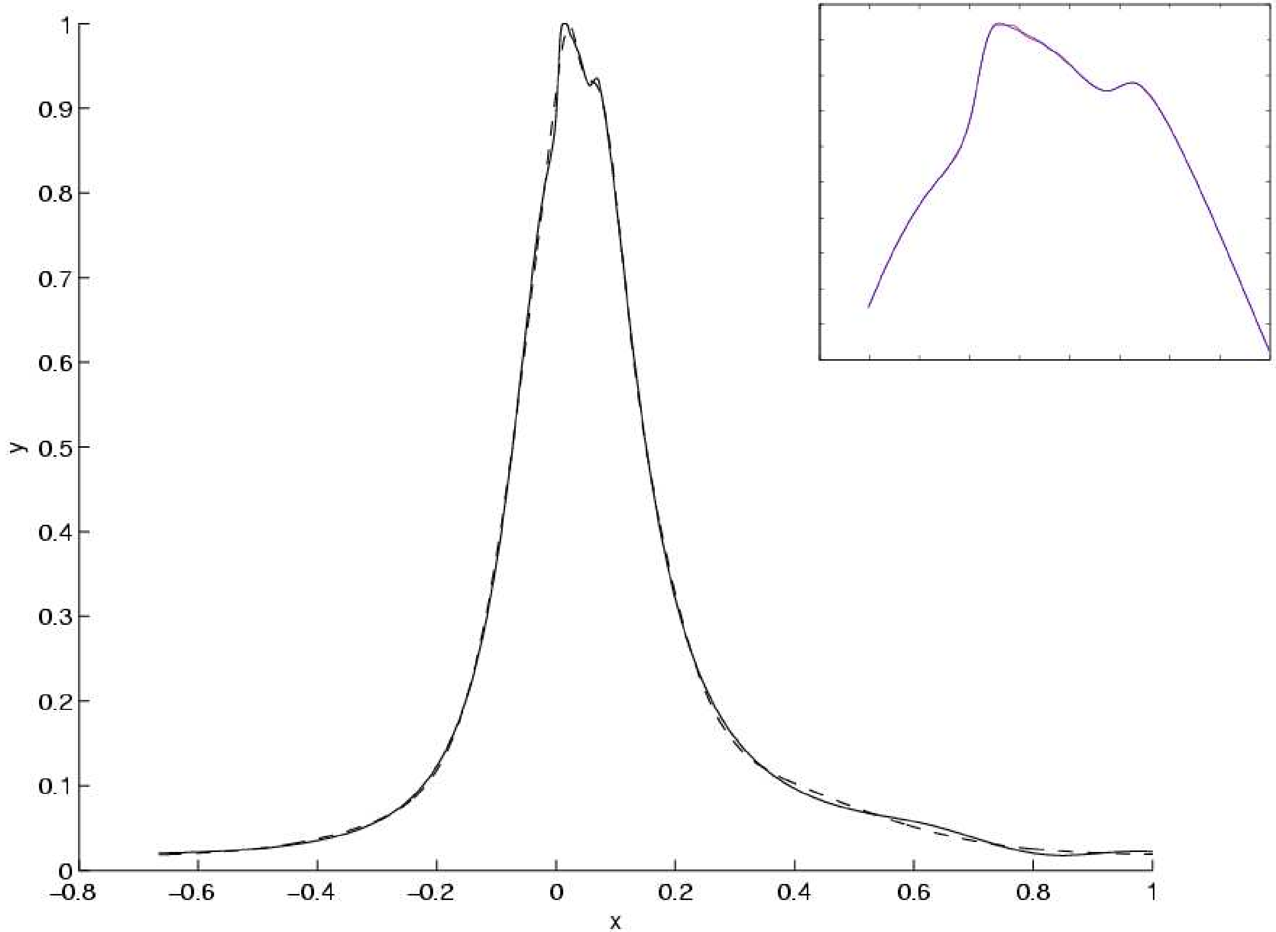
При селекции из объединенного множества родительских и порожденных функций в соответствии с критерием S выбираются M наилучших.

Конец итераций

Пример



Исходная и восстановленная выборка

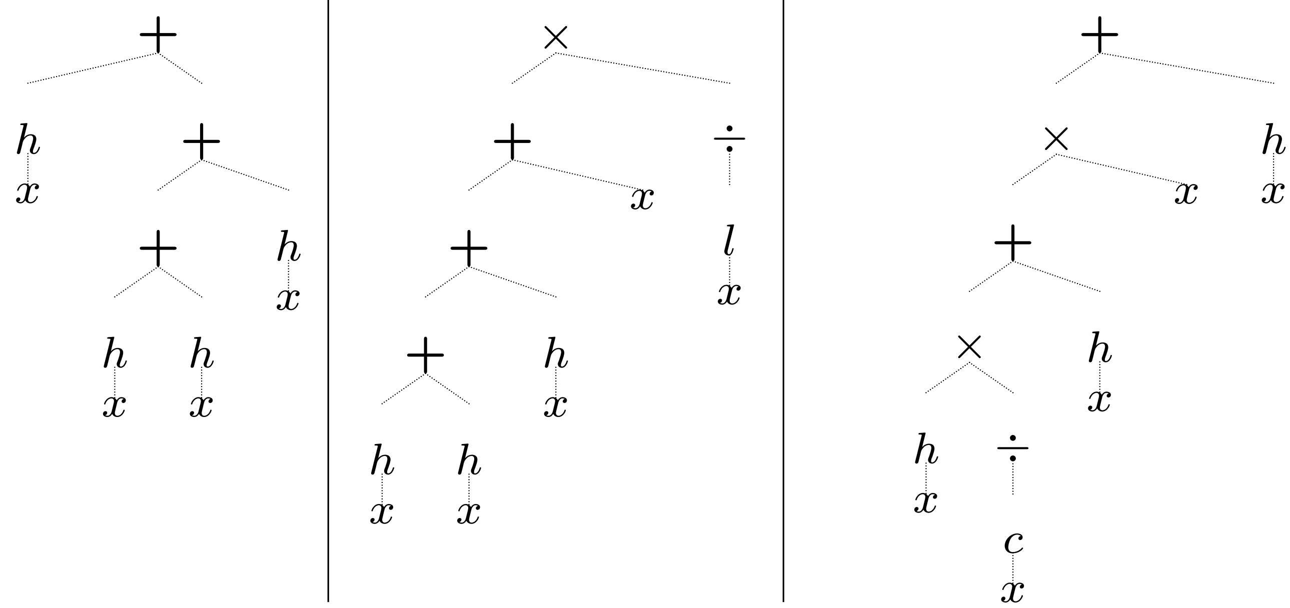


Порождающие функции

Функция	Описание	Параметры
	$g(\mathbf{b}, x_1, x_2)$	
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

Модели

f	$+(h, h, h, h)$	$\times(\div l, +(h, h, h, e))$	$+(\times(+(\times(h, \div h), h), e), h)$
mse$(y - \widehat{y})$	0.0034	0.0037	0.0035
max$(y - \widehat{y})$	0.0421	0.0325	0.00338
$\#\mathbf{w}$	16	16	16



Легенда: h — gaussian, c — cubic, l — logsig

Ответ

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

$$X \ni \mathbf{x},\ W \ni \mathbf{w}.$$

Выводы

Предложен способ получать несложные адекватные модели, которые возможно интерпретировать.

Поиск моделей в неявно заданном множестве $\mathcal{F}$ выполняется путем анализа гиперпараметров элементов моделей.