

ПОИСК РЕГРЕССИОННЫХ МОДЕЛЕЙ В ИНДУКТИВНО ЗАДАННОМ МНОЖЕСТВЕ

Стрижов В.В.

Вычислительный центр РАН

e-mail: strijov@ccas.ru

На заданной выборке значений нескольких свободных и одной зависимой переменной и заданном наборе непорождаемых гладких функций, обобщенно-индуктивно определяющих множество нелинейных регрессионных моделей рассмотрена процедура поиска регрессионной модели, оптимальной в евклидовой метрике.

Задана выборка $D = (\{\mathbf{x}_1, \dots, \mathbf{x}_N | \mathbf{x} \in \mathbb{R}^M\}, \{y_1, \dots, y_N | y \in \mathbb{R}\})$. Задано множество параметрических гладких функций $\mathcal{G} = \{g | g : \mathbb{R} \times \dots \times \mathbb{R} \rightarrow \mathbb{R}\}$, $g = g(\mathbf{b}, \cdot, \cdot, \dots, \cdot)$, где первый аргумент — параметр, а последующие — преобразованные свободные переменные. \mathcal{G} порождает множество суперпозиций $\mathcal{F} = \{f_i(\mathbf{w}, \mathbf{x})\}$ путем обобщенного индуктивного определения. Вектор параметров $\mathbf{w} = \mathbf{b}_1 : \mathbf{b}_2 : \dots : \mathbf{b}_r$, $\mathbf{w} \in \mathbb{R}^W$, где $:$ — знак присоединения векторов. Требуется найти регрессионную модель f_i , которая доставляет максимум функции плотности распределения $p(\mathbf{w} | D, \alpha, \beta, f_i)$, где α и β — регуляризующие параметры [1].

Предложен итерационный алгоритм поиска модели $y = f_i(\mathbf{w}, \mathbf{x}) + \nu$ с аддитивным гауссовским шумом. Алгоритм включает генетические оптимизационные процедуры поиска неявно заданных моделей [2] и метод доверительных интервалов для настройки их параметров. Работа поддержана грантом РФФИ 04-01-00401.

СПИСОК ЛИТЕРАТУРЫ

1. *MacKay, D.* Hyperparameters: optimise or integrate out? // Heidberger, G., ed. Maximum entropy and Bayesian Methods. Dordrecht: Kluwer, 1993.
2. *LeCun, Y., Denker, J. S., and Solla, S. A.* Optimal brain damage // Touretzky, D.S., ed., Advances in Neural Information Processing Systems. Morgan Kaufmann, San Mateo, CA, 1990. P. 598–605.