

Секция «Математика и механика»

Устойчивость существенно нелинейных регрессионных моделей и метод её исследования

Рудой Георгий Игоревич

Студент

Московский физико-технический институт, Факультет управления и прикладной математики, Москва, Россия

E-mail: 0xd34df00d@gmail.com

Анализ физического эксперимента состоит в восстановлении экспертно интерпретируемой функциональной зависимости между измеряемыми величинами. Одним из методов, позволяющих строить структурно сложные интерпретируемые модели, является символьная регрессия. Различные приближения сравниваются согласно некоторому функционалу ошибки, оптимизация параметров модели проводится с помощью, например, алгоритма Левенберга-Марквардта.

Принципиально важное значение имеют не только сами параметры функциональной зависимости, но и погрешность их определения, обусловленная погрешностями измеряемых величин. Для линейной регрессии дисперсия параметров известна для случая, когда регрессор известен точно, а дисперсия зависимой переменной во всех экспериментальных точках одинакова. Для случая нелинейной регрессии, а также ситуации, когда необходимо учитывать погрешности как регрессора, так и зависимой переменной, подобная задача, насколько нам известно, не ставилась.

В настоящей работе предложен метод решения этой задачи.

Сначала для данной выборки среди некоторого индуктивно порождаемого множества моделей находится параметрическая формула f , минимизирующая функционал качества, учитывающий, например, среднеквадратичную ошибку и сложность формулы [1].

Затем фиксируется структурный вид f и генерируется «достаточно большое» число реализаций значений нормально распределенной векторной случайной величины ξ (ее размерность равно количеству переменных в обучающей выборке) с заданной дисперсией, которое ω_i эмпирическое стандартное отклонение $D\omega_i$.

Процедура повторяется для различных значений дисперсии зашумляющей случайной величины $D\xi$, что позволяет определить зависимость стандартного отклонения параметров от $D\xi$ (компоненты признаков описания) $\frac{1}{D}\xi_i \{ \bar{x}_i \}$. Устойчивость модели f — это максимум устойчивости.

В численном эксперименте предложенный метод успешно применен для восстановления зависимости показателя преломления полимера n от длины волны λ . Оказалось, что для более корректной с физической точки зрения модели устойчивость «лучше», а скорость ее изменения при увеличении ошибок измерения n и λ меньше, чем для некорректных моделей.

Устойчивость регрессионной модели возможно использовать как критерий выбора оптимальной модели наряду с такими параметрами, как структурная сложность и функционал качества.

Литература

1. Рудой Г.И., Стрижов В.В. Алгоритмы индуктивного порождения суперпозиций для аппроксимации измеряемых данных // Информатика и ее применения. М, 2013. №. 7. С. 44-53.