

## Исследование и выбор регрессионных моделей с учетом погрешности как зависимых, так и независимых переменных

*Рудой Георгий Игоревич*<sup>1</sup>

0xd34df00d@gmail.com

<sup>1</sup>Москва, Московский физико-технический институт

В ходе анализа результатов физического эксперимента требуется восстановить функциональную зависимость между измеряемыми величинами, при этом необходима возможность экспертной интерпретируемости соответствующей зависимости.

Рассматривается класс существенно нелинейных параметрических суперпозиций элементарных функций, множество которых может быть либо заранее предложено экспертами, либо порождено, например, методом символьной регрессии.

Однако в ходе физического эксперимента измеряемые величины, как независимые, так и зависимые, известны лишь с некоторой конечной точностью. Соответственно, требуется не только выбрать регрессионную модель, минимизирующую сумму квадратов регрессионных остатков, но и оценить зависимость вариации ее параметров от вариации входных данных в рамках некоторых предположений о погрешностях измерения.

Кроме того, непосредственно в процессе построения модели необходимо учитывать погрешности измерения независимых переменных, для чего стандартный функционал среднеквадратичной ошибки не подходит.

В настоящей работе предлагается критерий устойчивости моделей, описывающий зависимость вариации параметров модели от вариации обучающих данных, и вводится функционал ошибки, учитывающий погрешности измерения независимых переменных, и соответствующий алгоритм оптимизации, основанный на алгоритме Левенберга-Марквардта. Предложенный метод также может быть применен в случае различия погрешности определения каждой из физических величин в различных точках.

- [1] *Рудой Г. И.* О возможности применения методов Монте-Карло в анализе нелинейных регрессионных моделей // Сибирский Журнал Вычислительной Математики, Новосибирск: Springer, 2015. — *принято в печать*

## Model selection and analysis with respect to measurement errors in both free and dependent variables

Georg Rudoy<sup>1</sup>

0xd34df00d@gmail.com

<sup>1</sup>Moscow, MIPT

A functional dependency between measured data is to be found during physical experiment analysis. It is desirable for the dependency to be interpretable by experts.

For this, a set of non-linear parametric superpositions of elementary functions is considered (either inductively generated via, for example, symbolic regression, or proposed by experts) as the set of candidate regression models.

The data is measured with some finite precision during the physical experiment, though. Thus, it is also required to estimate the dependency of variation of the selected model parameters on the variation of the learning data, accounting for measurement errors.

Moreover, during the model selection process (and, in particular, during parameters optimization) the possible measurement errors of independent variables should be taken into account. The standard MSE functional does not fit this task since it assumes zero measurement error of these variables.

This work proposes a model selection criterion called *model stability*, describing the dependency of model parameters variation on learning data variation, and a functional accounting for measurement errors in both dependent and independent variables along with the optimization method based on the Levenberg-Marquardt algorithm. The proposed criterion and method may also be used in case of different measurement errors in various data points.

- [1] Rudoy G. On applying Monte-Carlo methods to analysis of non-linear regression models // Numerical Analysis and Applications, Novosibirsk: Springer, 2015. — *accepted for printing*