Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 8, n. 1, 2021.

Trabalho apresentado no XL CNMAC, Evento Virtual - Co-organizado pela Universidade do Mato Grosso do Sul (UFMS).

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Relating Least Squares method input error with output error

Giovanni Guarnieri Soares[1]
Instituto Nacional de Pesquisas Espaciais (INPE), São José Dos Campos, SP
Leonardo Santos[2]
Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), São José Dos Campos, SP

The least squares method is the standard numerical method procedure for finding the best-fitting curve to a set of points, adjusting a curve looking to minimize the sum of the squares of the offsets (hence why least squares). These residuals of the points from the curve sometimes can have outlying points which can cause a disproportionate effect on the fit, which can be undesirable [2]. That's this works focus.

First we're going to generate a random angular coefficient with a normal distribution, from 0.0 to 100.0. This number is used in the equation:

$$y = ax + b + error. \tag{1}$$

To generate a dataset without offsets. Then with a random number generator we're going to introduce these offsets, ranging from 0% to 100% of the actual data.

By applying artificial offsets in a randomly generated line we're going to show the relationship of the data error with the fit error. These offsets are introduced via a random number generator routine using two different distributions, uniform and normal (Gaussian) [1].

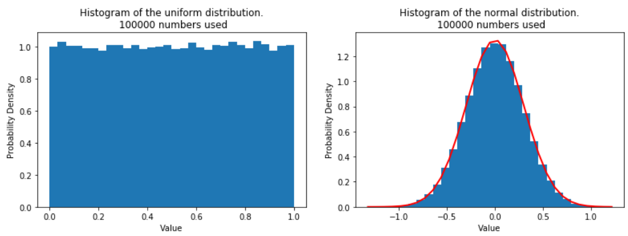

Figure 1: Graphs illustrating both distributions. First an uniform distribution generating values from 0 to 1 and second a normal distribution with $\sigma = 0.3$ and $\mu = 0$.

These datasets are going to be fitted by two steps coming from the same method of least squares, being them

$$a = \frac{(\sum_{i=1}^{n} x_i y_i) - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}; \tag{2}$$

[1]giovanniguarnieri@id.uff.br.
[2]santoslbl@gmail.com.

and

$$a = \frac{cov(x, y)}{\sigma_x^2}, \tag{3}$$

Which are shown next, respectively for each distribution either

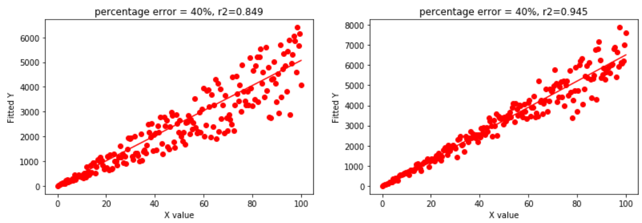

Figure 2: Graphs showing the linear regression made with least squares using the equation 2, errors added with uniform and normal distribution respectively.

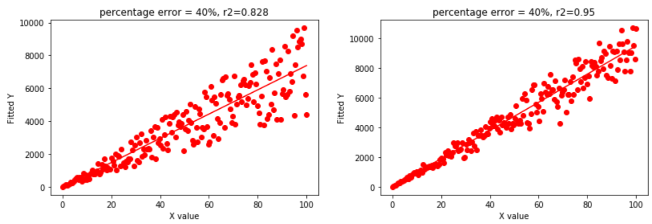

Figure 3: Graphs showing the linear regression made with least squares using the equation 3, errors added with uniform and normal distribution respectively.

All of them are going to show promising results, with maximum 6% percentage error for the output to a noise with the possibility of the same size of the data input.

Acknowledge

Funding:
São Paulo Research Foundation (FAPESP), Grant Number 2015/50122-0 and
DFG-IRTG Grant Number 1740/2;
FAPESP Grant Number 2018/06205-7;
CNPq Grant Number 420338/2018-7.

References

[1] E. Çinlar. *Probability and Stochastics.* Springer New York, 2011.

[2] E. W. Weisstein. Least squares fitting, 2018. [Online; accessed 20/01/2020].