

A robust technique to transform time series with missing data into a zero mean series

K.K.L.B. Adikaram^{1*} and P.A. Jayantha²

¹*Computer Unit, Faculty of Agriculture, University of Ruhuna, Sri Lanka,*

²*Department of Mathematic, Faculty of Science, University of Ruhuna, Sri Lanka*

^{*}*lasantha@agricc.ruh.ac.lk*

In the process of knowledge mining, feature scaling plays a crucial role. For feature scaling, there are two widely used methods in many machine learning algorithms, namely, mean normalization and standardization. In both these methods, mean is zero and consider only the dependent variable (y); no involvement of independent variable (x). Thus, regardless of the original regression, those approaches treat all the data as $y = c$ series. However, when some data points are missing or removed (as outliers), the said approaches destroy the original regression. In this research, a novel standardization method to standardize data based on linear regression ($y = mx + c$) was introduced. The proposed transformation was given by $y_{new} = y^T - x \times \overline{y^T} / \overline{x^T} + c$, where $y^T = y - y_r$, $x^T = x - x_r$, (x_r, y_r) is any selected reference point, and c is any constant. The speciality of the proposed method is that the transformation used both independent and dependent variables. Thus, no influences from missing or removed data. When there were no noise or outliers, the new method transformed data into a $y = c$ series even with multiple missing values. When $c = 0$, results showed that the transformation produces a zero mean series from any data set even with noise, outliers and missing values.