Department of Computer Science

St. Francis Xavier University

Presents

Novel Metrics for Evaluation and Validation of Regression-Based Supervised Learning

by

Md. Mostafizur Rahman

St. Francis Xavier University

M.Sc. Thesis Proposal Presentation

February 4, 2022 at 9am AST via Zoom

Error consistency is a validation metric for evaluating the sample-based error variability across machine learning models trained as part of in-lab validation. Many machine learning (ML) based regression algorithms are likely to be inconsistent when trained repeatedly on the same task, in part due to sampling, but also, potentially associated with the inclusion of randomness in their training paradigms, which is common in many learning techniques. In this work, we propose a novel approach to validation and evaluation of regression-based learning algorithms, called 'regression error consistency (EC)' to assist in assessing sample-wise consistency of errors as part of in-lab validation. We have applied our proposed novel EC metrics on simulated datasets generated from different distributions, including outliers and heteroscedasticity properties with three regression algorithms (i.e, Knn-1, Knn-5, and Linear regression) and found a correlation with well-known goodness of fit (GoF) metrics like mean absolute error (MAE), mean squared error (MSE), and R-Squared (R2).We have also used EC metrics on seven real-world datasets with six different regression algorithms and compared the results with classification EC (from a previously completed MSc thesis). The result demonstrated that the random forest (out of six models used in the experiment) achieved high accuracy but showed less consistency in its error profiles. This finding matches with classification based EC results. In addition, we compare the correlation between classification EC and regression EC using three real-world datasets and four ML models. The findings demonstrate similar behaviour between the classifiers and the regression learners. Moreover, we applied the EC metric on the MNIST digits dataset using a convolutional neural network (CNN) as part of our deep learning experiment. Though MNIST is typically treated as a classification dataset, we considered this dataset as a regression problem and the CNN model developed demonstrated good performance. We believe that the proposed EC metrics will be useful in evaluating and validating regression algorithm error consistency, including in deep learning, and will hopefully guide the machine learning research community to develop more reproducible and predictable (in terms of the errors they will make) regression algorithms. Future work will extend our regression error consistency validation methods to two advanced biomedical applications of deep learning, one long short term memory (LSTM) recurrent neural network (RNN) on a biomedical public dataset for at risk patient monitoring, and one regression UNet for longitudinal image estimation of patients being monitored for neural degeneration. Moreover, we will implement an additional EC metric based on intersections and unions of the residual error bars.