

Representativity Based Wasserstein Active Regression

Authors : Benjamin Bobbia, Matthias Picard

Abstract : In recent years active learning methodologies based on the representativity of the data seems more promising to limit overfitting. The presented query methodology for regression using the Wasserstein distance measuring the representativity of our labelled dataset compared to the global distribution. In this work a crucial use of GroupSort Neural Networks is made therewith to draw a double advantage. The Wasserstein distance can be exactly expressed in terms of such neural networks. Moreover, one can provide explicit bounds for their size and depth together with rates of convergence. However, heterogeneity of the dataset is also considered by weighting the Wasserstein distance with the error of approximation at the previous step of active learning. Such an approach leads to a reduction of overfitting and high prediction performance after few steps of query. After having detailed the methodology and algorithm, an empirical study is presented in order to investigate the range of our hyperparameters. The performances of this method are compared, in terms of numbers of query needed, with other classical and recent query methods on several UCI datasets.

Keywords : active learning, Lipschitz regularization, neural networks, optimal transport, regression

Conference Title : ICAL 2023 : International Conference on Active Learning

Conference Location : Paris, France

Conference Dates : October 30-31, 2023