

A GENERAL FRAMEWORK FOR ACTIVE LEARNING IN REGRESSION, WITH APPLICATIONS TO NUMERICAL PDES

BEN ADCOCK, JUAN M. CARDENAS, AND NICK DEXTER

ABSTRACT. Active learning is an important concept in machine learning, in which the learning algorithm is able to choose where to query the underlying ground truth to improve the accuracy of the learned model. As machine learning techniques come to be more commonly used in scientific computing problems, where data is often expensive to obtain, the use of active learning is expected to be particularly important in the design of efficient algorithms. In this work, we introduce a general framework for active learning in regression problems. Our framework extends the standard setup by allowing for general types of data, rather than merely pointwise samples of the target function. This generalization covers many cases of practical interest, such as data acquired in transform domains (e.g., Fourier data), vector-valued data (e.g., gradient-augmented data), data acquired along continuous curves, and, multimodal data (i.e., combinations of different types of measurements). Our framework considers random sampling according to a finite number of sampling measures and arbitrary nonlinear approximation spaces (model classes). We introduce the concept of generalized Christoffel functions and show how these can be used to optimize the sampling measures. We prove that this leads to near-optimal sample complexity in various important cases. This work focuses on applications in scientific computing, where, as noted, active learning is often desirable, since it is usually expensive to generate data. We demonstrate the use of these ideas on several different scientific computing problems, including learning the solution map of high-dimensional parametric PDEs and the numerical solution of PDEs via Physics-Informed Neural Networks (PINNs). For the latter, we show how active learning can be deployed in an adaptive manner to specifically target the sampling scheme to the behaviour of the unknown PDE solution.

Keywords: active learning, leverage scores, Christoffel functions, parametric PDEs, Physics-Informed Neural Networks (PINNs)

Mathematics Subject Classifications (2010): 65D40, 65N50, 68T07, 68Q32

REFERENCES

- [1] B. Adcock, J. M. Cardenas and N. Dexter *CS4ML: A general framework for active learning with arbitrary data based on Christoffel functions*. *arXiv:2306.00945*, 2023.
- [2] B. Adcock, J. M. Cardenas and N. Dexter CAS4DL: Christoffel adaptive sampling for function approximation via deep learning. *Sampling Theory, Signal Processing, and Data Analysis*, 20:2(21), 2022.
- [3] B. Adcock and J. M. Cardenas Near-optimal sampling strategies for multivariate function approximation on general domains. *SIAM Journal on Mathematics of Data Science*, 2(3):607-630, 2020.

SIMON FRASER UNIVERSITY
Email address: ben_adcock@sfu.ca

UNIVERSITY OF COLORADO, BOULDER
Email address: juan_manuel_cardenas@sfu.ca

FLORIDA STATE UNIVERSITY
Email address: nick.dexter@fsu.edu