Dan D. Padilha
QxBranch, LLC

Training L0-Regularised Linear Regression and Classification Models with a Quantum Annealer

Feature selection is a critical element of many applications of machine learning models. By selecting an optimal subset of features (variables) that best explain the underlying realities of a given dataset, a model becomes simpler to understand, more computationally efficient, and less susceptible to overfitting and noise. For this project, QxBranch designed, implemented, and benchmarked quantum-based $L0$-regularised (Q$L0$) linear regression and classification models on real-world and synthetic datasets. The $L0$ regularisation, a form of feature selection which penalises selecting too many features for a given model, encourages greater sparsity than other classical regularisation methods, but is a computationally intractable problem. Our Q$L0$ implementation used the DW2X hardware to demonstrate generalised configurable-precision predictive models with linear combinations of selected features. The hybrid classical-quantum algorithm for Q$L0$ was implemented as a scikit-learn (commonly-used Python machine learning library) "estimator" interface, allowing for simple integration with existing machine learning and scientific analysis pipelines. Our research showed that across the tested regression and classification datasets, Q$L0$ consistently resulted in high-accuracy sparse prediction models that selected sets of features approximating those selected by leading sparse non-linear models. Based on these results, it is possible that quantum-based $L0$ regularisation methods may perform well as sparse feature selection steps in many machine learning algorithms.