

CLASSICAL DATA PREPROCESSING INFLUENCE ON QML

**SOHAM PAL, DANIEL SIERRA-SOSA (MENTOR), MICHAEL TELAHUN,
IQRA NAZ**

Created: 2021-04-01 Thu 14:38

WHAT IT IS ABOUT?

Evaluation of the suitability of real-world datasets for processing with Quantum Machine Learning and Quantum Assisted Machine Learning.

- Case study 1
 - **Synthetic data generation for QML.** Generate data from various statistical distributions for classification problems.
 - **Benchmark QML algorithms.** We will benchmark how the out-of-the-box QML algorithms, included in Qiskit and PennyLane, like QSVM and VQC perform on these datasets. We also want to see the effects of classical pre-processing, such as rotation, scaling, etc impacts the QML algorithms.
- Case study 2
 - We want to use a variants of anle encoding and amplitude encoding for the quantum representation of the synthetic data, and if these make the QML algorithms data-agnostic, and what are the effects of classical pre-processing.
 - We will use this encoded data to benchmark performance of different frameworks for QML - particularly Qiskit + Pytorch, Tensorflow Quantum, and PennyLane.
- Case study 3 (if time permits)
 - Do the above with a real dataset.

WHAT WE HAVE DONE AND GOING TO DO?

- Progress
 - Developed generators for synthetic datasets from multiple statistical distributions.
 - Studying the effects of rotation on datasets from the normal distribution. As it appears rotating the data can help to improve the performance of QSVM and VQC.
- Deliverables
 - Journal submissions
 - Benchmark of out-of-the-box QML algorithms with synthetic datasets.
 - Evaluation of mapping strategies for particular case studies.