Quantum Machine Learning Classifiers

Bollu Venkata Adithya **24116025**

1. Objective

The primary objective of this project is to design, implement, and benchmark a Hybrid Quantum-Classical Classifier for noisy, non-convex classification tasks. My goal is to evaluate if Quantum Machine Learning (QML) models, specifically Variational Quantum Circuits (VQCs), can offer competitive performance or advantages over classical baselines (Logistic Regression, Random Forest) on high-dimensional datasets when subjected to dimensionality reduction.

2. Overview

This project explores the application of QML on two distinct datasets:
1. Credit Card Fraud Detection: A highly imbalanced, high-volume dataset.
2. Heart Disease Prediction: A medical diagnostic dataset with complex feature interactions.

I utilized PennyLane for quantum circuit simulation and PyTorch for constructing the hybrid neural network architecture. The analysis includes data preprocessing, classical benchmarking, hybrid model training, and a final comparative assessment using Area Under the ROC Curve (AUC) and Accuracy metrics. A bonus investigation into Data Re-uploading Quantum Neural Networks (QNNs) was also conducted.

3. Data Analysis and Preprocessing

To adapt the classical datasets for a 4-qubit quantum simulation, I applied the following pipeline:

Fraud Detection Dataset:
- Original Dimensions: 8 features.
- Cleaning: I removed rows with missing values.
- Scaling: I applied StandardScaler for zero mean and unit variance.
- Dimensionality Reduction: I utilized PCA to compress 8 features into 4 principal components, retaining ~59% of the variance.

Heart Disease Dataset:
- Original Dimensions: 13 features.
- I reduced 13 features to 4 components via PCA (retaining ~51% variance). This aggressive reduction posed a challenge for feature retention.

4. Quantum Circuit Design

The core of my solution is a Variational Quantum Circuit (VQC) with two main blocks:

1. Feature Map (Encoding):
- Type: AngleEmbedding
- Function: Encodes the 4 classical features into the rotation angles of qubits (e.g. Rx), mapping classical data to a quantum state.

2. Ansatz (Variational Layers):
- Type: StronglyEntanglingLayers
- Depth: 2 layers.
- Function: Applies a series of parameterized rotations and CNOT entangling gates optimized during training.

(Bonus) Data Re-uploading QNN:
I implemented a pure QNN where the encoding and variational layers are interleaved (repeated 4 times) to increase expressivity.

5. Training Pipeline

I established a Hybrid Quantum-Classical training loop:

- Framework: PyTorch (qml.qnn.TorchLayer).
- Quantum Layer: 4-qubit VQC outputting 4 expectation values.
- Classical Post-processing: A Linear Layer mapping 4 quantum outputs to 1 prediction, followed by Sigmoid activation.
- Optimizer: Adam (Learning Rate = 0.01).
- Loss Function: Binary Cross Entropy (nn.BCELoss).

6. Comparative Analysis

Dataset 1: Fraud Detection

Model	Accuracy	AUC
Logistic Regression	91.35%	0.8812
Random Forest	99.85%	0.9999
Hybrid QML (Proposed)	97.03%	0.9678

Analysis: The Hybrid QML model significantly outperformed the linear baseline (Logistic Regression) and achieved a high AUC (>0.96).

Dataset 2: Heart Disease

Model	Accuracy	AUC
Logistic Regression	83.90%	0.9253
Random Forest	100.00%	1.0000
Hybrid QML	67.32%	0.7810

Analysis: The QML model underperformed here, primarily due to aggressive PCA reduction (13 -> 4 features) losing information.

7. Conclusion

This project successfully demonstrated the end-to-end implementation of a Hybrid QML classifier.
1. Success: On the Fraud Detection dataset, the QML model proved highly effective.
2. Challenge: The Heart Disease analysis highlighted the critical dependency of NISQ-era QML on effective dimensionality reduction.
3. Future: The Bonus Data Re-uploading architecture suggests a path forward to improve expressivity without increasing qubit count.