Quantum Machine Learnining

Julius Ruseckas

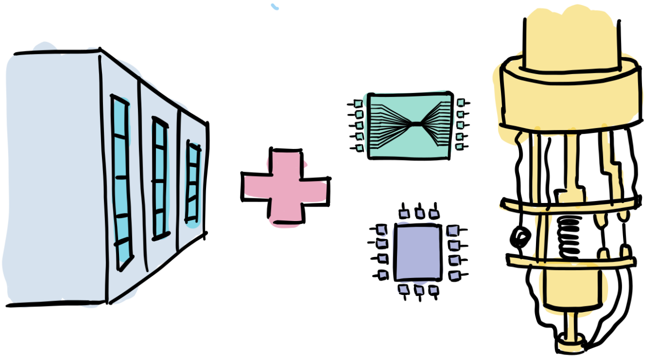
July 1, 2022

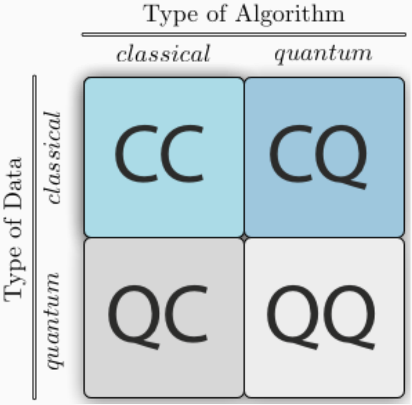
Baltic Institute of Advanced Technology

Outline

1. What is Quantum Machine Learning
2. Data input
3. Classifier using parametrized quantum circuits
4. Problems
5. Summary

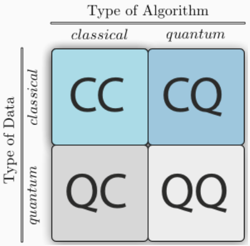
What is Quantum Machine Learning





Different approaches to combine quantum computing and machine learning

- Quantum Machine Learning uses a quantum device for performing machine learning tasks.
- Quantum Machine Learning uses a quantum device for classification or feature extraction from quantum states.
- Larger speed or larger accuracy than it is possible using classical machine learning methods.



There are known algorithms for a quantum computer that perform faster than classical counterparts:

- HHL algorithm for solving of a system of linear equations.
    - Performs a matrix inversion using an amount of resources growing only logarithmically in the dimensions of the matrix.
    - Requirement: matrix should be sparse or low rank
- Quantum Fourier transform
- Amplitude amplification methods based on Grover's search algorithm

- Quantum k-means clustering
- Quantum SVM
- Quantum PCA
- Quantum neural networks

Data input

When QML algorithm processes classical data, one needs to encode classical data to a quantum state. Two encoding methods:

- Bit encoding
- Amplitude encoding

Bit encoding

i-th data record consisting of a squence of N bits $\mathbf{b}^{(i)} = (b_1, b_2, \dots b_N)$ is encoded as a sequence of qubits

$$O|i\rangle|0\rangle = |i\rangle|\mathbf{b}^{(i)}\rangle \, ,$$

where O is an operator representing an data oracle.

Bit encoding is used in

- Quantum SVM
- Quantum k-nearest neighbours

Bit encoding

- Advantage: data is presnted in the same way as in classical ML algorithm
- Disadvantage: needs large number of qubits

Data record $x = (x_1, x_2, \ldots x_N)$ having N elements is encoded in superposition amplitudes

$$|\varphi(x)\rangle = \frac{1}{\chi}\sum_{i=1}^{N} x_i|i\rangle \,, \qquad \chi = \sqrt{\sum_{i=1}^{N} x_i^2}\,.$$

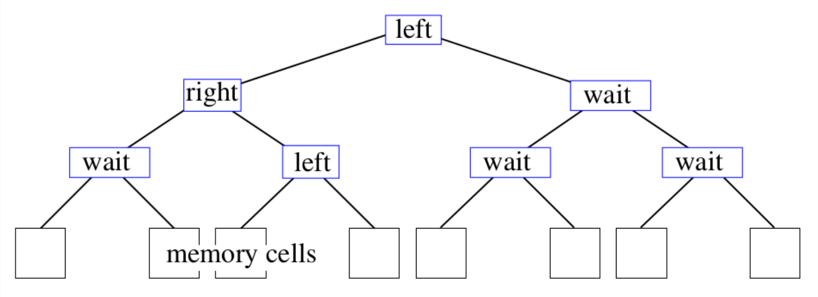
Here $|i\rangle$ are basis vectors.

Amplitude encoding is used in

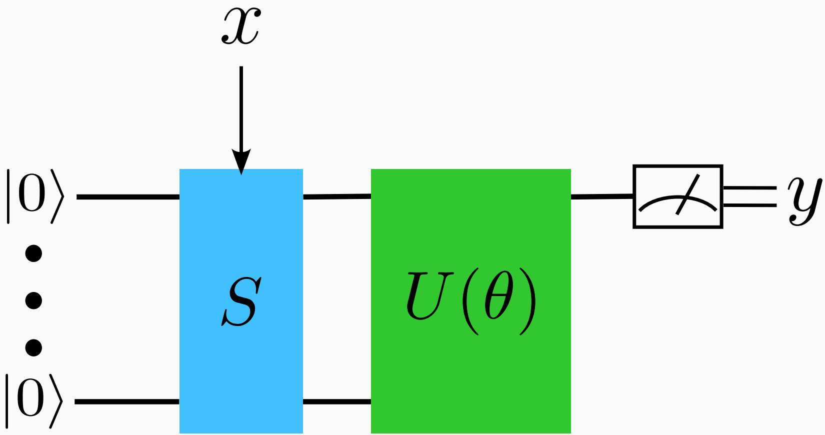
- Quantum k-nearest neighbours
- Quantum neural networks

Advantage: for encoding N numbers one needs $n = \log_2 N$ qubits. Exponentially compact representation.

- qRAM uses n qubits to address any quantum superposition of $N = 2^n$ memory cells.
- $O(\log N)$ switches are needed to select memory cell

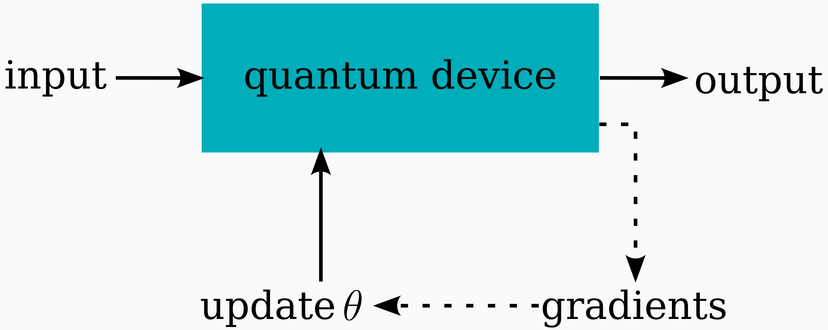


Classifier using parametrized quantum circuits


x
$|0\rangle$
$|0\rangle$
S
$U(\theta)$
y

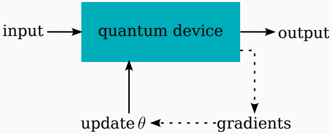
Classifier using parametrized quantum circuits

- Quantum circuit parametrized by parameters θ
- Represented by an unitary operator $U(\theta)$
- Input data x are represented by a quantum state $|\varphi(x)\rangle$
- Quantum circuit transforms the input state to $U(\theta)|\varphi(x)\rangle$
- Classification result $f(x, \theta)$ is measured on the output qubit in the state $U(\theta)|\varphi(x)\rangle$
- Parameters θ are optimized to get correct classification results


input
quantum device
output
update θ
gradients

Training algorithm

- Quantum circuit parameters θ are optimized using classical gradient descent algorithm
- Derivatives $\partial_\theta U(\theta)$ with respect to the parameters are calculated by a quantum circuit



Amplitude encoding is used for input

$$|\varphi(x)\rangle = \frac{1}{\chi}\sum_{i=1}^{N} x_i|i\rangle\,, \qquad \chi = \sqrt{\sum_{i=1}^{N} x_i^2}\,.$$

using n qubits, $2^n = N$

Objective function

Predicted labels $\ell(x) = \lambda_1, \lambda_2, \ldots$ are obtained by measuring the output state. The measurement corresponds to a Hermitian operator

$$A = \sum_j \lambda_j P_{\lambda_j}\,,$$

where P_{λ_j} are projection operators to the eigenspace corresponding to λ_j.

Objective function: mean likelihood of inferring the correct label

$$\mathcal{L}(\theta) = \frac{1}{M} \sum_{x:\ell(x)=\lambda_j} \langle \varphi(x) | U^\dagger(\theta) P_{\lambda_j} U(\theta) | \varphi(x) \rangle\,,$$

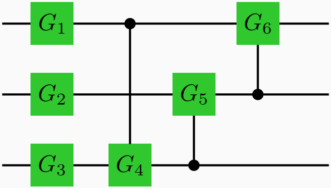
where M is the size of the dataset.

Structure of quantum circuit

- Quantum circuit are constructed from blocks

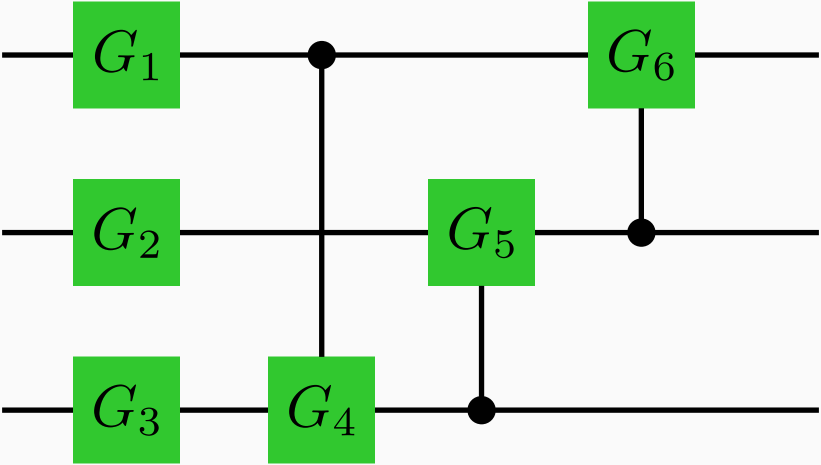
$$U(\theta) = G_{\mathrm{out}}(\theta_{\mathrm{out}}) B_L(\theta_L) \cdots B_2(\theta_2) B_1(\theta_1)$$

- Each block $B_l(\theta_l)$ consist of n single-qubit gates and n controlled single-qubit gates forming a cyclic code.



- Cyclic code is characterized by a hyperparameter $0 < r < n$.
- For any qubit with index j there should be one controlled single-qubit gate with the j-th qubit as the target and $(j + r) \mod n$-th qubit as the control.
- A cyclic code block can cause a near-maximum increase or decrease of the bipartite entanglement entropy.

Rapidly entangling block



Single-qubit gate

$$G(\alpha,\beta,\gamma) = \begin{pmatrix} \mathrm{e}^{\mathrm{i}\beta}\cos\alpha & \mathrm{e}^{\mathrm{i}\gamma}\sin\alpha \\ -\mathrm{e}^{-\mathrm{i}\gamma}\sin\alpha & \mathrm{e}^{-\mathrm{i}\beta}\cos\alpha \end{pmatrix}, \qquad \alpha,\beta,\gamma \in [0,\pi]$$

Controlled single-qubit gate

$$|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes G$$

Partial derivatives

- The partial derivatives of single-qubit gate $G(\alpha, \beta, \gamma)$ yield a gate of the same type.
- A partial derivative of a controlled single-qubit gate is no longer a unitary operator
- A partial derivative of a controlled single-qubit gate can be represented as a linear combination of two unitary operators

$$\partial_\theta \{|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes G\} = |1\rangle\langle 1| \otimes \partial_\theta G = \frac{1}{2}(I \otimes \partial_\theta G - \sigma_z \otimes \partial_\theta G)$$

- Partial derivatives $\partial_\theta U(\theta)$ can be represented as a certain linear combination of unitary circuits of the same architecture.
- The sum of the terms in the derivatives are calculated by classical postprocessing.

Problems

Problems

- Barren plateaus (exponentially vanishing gradients)
- Data loading can take a large number of steps

Barren plateaus

- Random initialization of parameters.
- The probability that the gradient along any direction is non-zero is exponentially small as a function of the number of qubits.
- Can be caused by noise in quantum device

Barren plateaus

- Exponential dimension of Hilbert space.
- Entanglement between the visible and hidden units in a quantum neural network
- Entanglement causes information to be non-locally stored in the correlations between the qubits rather than in the qubits themselves.
- Measuring only visible qubits, such information is lost.

Summary

Summary

- Quantum computers can accelerate ML algorithms
- For some ML tasks, even quantum computers with moderate number (~ 100) of qubits can be useful
- However, it is still unclear if QML provides practical advantage

Thank you for your attention!