Summary: An introduction to QML

1 Introduction

To discuss the various approaches to relate standard methods of machine learning to quantum physics.

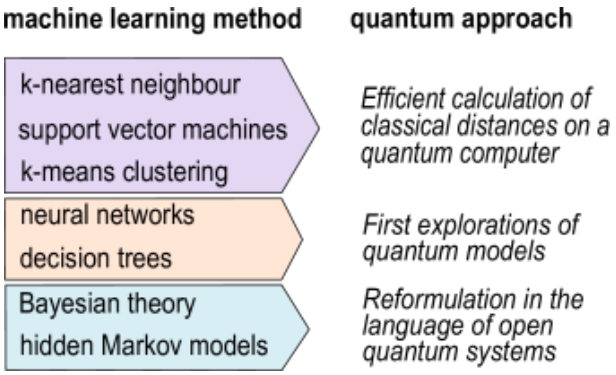

Figure 1: Overview of methods in machine learning and approaches from a quantum information perspective

In quantum machine learning, quantum algorithms are developed to solve typical problems of machine learning using the efficiency of quantum computing. This is usually done by adapting classical algorithms or their expensive subroutines to run on a potential quantum computer.

2 Quantum versions of ML algorithms

2.1 Quantum versions of k-nearest neighbour methods

The nearest-centroid algorithm works by assigning a new input data point to the class corresponding to the centroid (i.e., the mean vector) of the class with the smallest distance to the input. K-nearest neighbours is obviously based on a distance metric to evaluate the similarity of two feature vectors. Hence, translating this algorithm into a quantum version therefore focuses on the efficient evaluation of a classical distance through a quantum algorithm.

ABG algorithm: ABG introduced the idea of using the overlap or fidelity of two quantum states $|a\rangle$ and $|b\rangle$ as a 'similarity measure'. The overlap or fidelity between two quantum states is defined as the absolute value of the inner product between the states, which is given by $|\langle a\ ,\ b\rangle|$. The fidelity between

two quantum states can be obtained through a simple quantum routine known as a swap test. In the swap test, the two quantum states are first encoded into quantum registers, and then a controlled-SWAP gate is applied to the registers. The controlled-SWAP gate swaps the two quantum states if a control qubit is in the $|1\rangle$ state, and leaves them unchanged if the control qubit is in the $|0\rangle$ state. After the controlled-SWAP gate is applied, a measurement is performed on the control qubit. If the measurement outcome is $|1\rangle$, it indicates that the two quantum states were not identical, and the fidelity between them is zero. If the measurement outcome is $|0\rangle$, it indicates that the two quantum states were identical, and the fidelity between them is one. The fidelity between two quantum states can be used as a similarity measure in the ABG quantum k-NN algorithm to compare the input data point with the data points in the training set and determine the k nearest neighbors. The labels of the k nearest neighbors are then used to classify the input data point.

Lloyd, Mohseni, and Rebentrost proposed a method to retrieve the distance between two real-valued n-dimensional vectors $\vec{a}$ and $\vec{b}$ using a quantum measurement, based on the use of a swap test, which is a simple quantum routine that can be used to determine the fidelity or overlap between two quantum states. To retrieve the distance between the vectors a and b, define a quantum state $|\psi\rangle = \frac{1}{\sqrt{2}}(|0, a\rangle + |1, b\rangle)$ and a quantum state $|\phi\rangle = \frac{1}{\sqrt{Z}}(|\vec{a}|\,|0\rangle - |\vec{b}|\,|1\rangle)$, where $Z = |\vec{a}|^2 + |\vec{b}|^2$. The quantum state $|\psi\rangle$ encodes the vectors $\vec{a}$ and $\vec{b}$ into a quantum superposition, while the quantum state $|\phi\rangle$ encodes the lengths of the vectors $\vec{a}$ and $\vec{b}$ into the scalar product of the quantum state with itself. Next, performing a swap test between the quantum states $|\psi\rangle$ and $|\phi\rangle$, and measuring the fidelity or overlap between the states. The fidelity between the states is given by $|\langle\psi|\phi\rangle|^2$, and the identity $|\vec{a} - \vec{b}|^2 = Z|\langle\psi|\phi\rangle|^2$ holds true, so the distance between the vectors a and b can be obtained from the fidelity.

In the quantum nearest-centroid algorithm, the vector a is taken to be the input data point, and the vector b is taken to be the mean vector of a particular class, which is given by $\frac{1}{N_c}\sum_p \vec{v_p}$. The distance between the input data point and the mean vector of the class is calculated using the swap test, and the input data point is assigned to the class corresponding to the mean vector with the smallest distance. Claim: even when considering the operations to construct the quantum states involved, this quantum method is more efficient than the polynomial runtime needed to calculate the same value on a classical computer.

Wiebe, Kapoor and Svore also use a swap test in order to calculate the inner product of two vectors, which is another distance measure between feature vectors.

2.2 Quantum computing for support vector machines

SVMs are based on the concept of finding the hyperplane in a high-dimensional space that maximally separates the data points of different classes. A support vector machine is used for linear discrimination, where the task is to find a hyperplane that is the best discrimination between two class regions and serves as a decision boundary for future classification tasks. It seems like a severe restric-

tion that methods of linear discrimination require the problem to be linearly separable, which means that there is a hyperplane that divides the datapoints so that all vectors of either class are on one side of the hyperplane (in other words, the regions of each class have to be disjunct). However, a non-separable problem can be mapped onto a linearly separable problem by increasing the dimensions.

This is a mathematical optimisation problem of finding the maximum margin $|\vec{w}|^{-1}(\vec{v}\vec{w}+b)$ between the hyperplane and the support vectors. Mathematical formulation of the optimisation problem contains a kernel K, a matrix containing the inner product of the feature vectors. $(K)_{pk} = \vec{v_p}\cdot\vec{v_k}$; p,k=1,...,N (or the basis vectors they are composed of) as entries. Support vector machines are in fact part of a larger class of so called kernel methods that suffer from the fact that calculating kernels can get very expensive in terms of computational resources. One of the key challenges in using SVMs is the calculation of the inner product between the data points, which is used to determine the similarity between the points. It is thus crucial for support vector machines to find a method of evaluating an inner product efficiently. This is where quantum computing comes into play.

In the quantum kernel method, the inner product between two data points is calculated using a quantum circuit consisting of a series of quantum gates applied to the data points. The quantum circuit is designed to encode the inner product into the quantum state of the output, which can be measured to obtain the inner product. Rebentrost, Mohseni, and Lloyd proposed the use of a quantum algorithm to calculate the inner product between data points. To apply the quantum algorithm to a set of training data, represent the training data as a 2^n-dimensional basis of the training vector space T, where the $\left|x^i\right\rangle$ are the basis vectors and every training vector $|v_p\rangle$ can be represented as a superposition $|v_p\rangle = \sum \alpha_i \left|x^i\right\rangle$. The kernel matrix of the inner products of the basis vectors,K with $(K)_{i,j} = \vec{x^i}\cdot\vec{x^j}$ can then be calculated by taking the partial trace of the corresponding density matrix.

2.3 Quantum algorithms for clustering

The **Aimeur, Brassard, Gilles and Gambs** quantum k-median algorithm is a quantum algorithm for clustering data points based on their distance from a set of medians. The algorithm uses two subroutines to achieve this: a distance calculation subroutine and a find minimum subroutine. The distance calculation subroutine calculates the distance between two quantum states using an oracle. An oracle is a quantum subroutine that can perform a specific computation, such as calculating the distance between two quantum states. The distance calculation subroutine calculates the total distance of each state to all other states of one cluster. The find minimum subroutine is used to find the smallest value of the distance function calculated by the distance calculation subroutine. The find minimum subroutine selects the quantum state corresponding to the smallest distance value as the new median for the cluster.

The **Lloyd, Mohseni and Rebentrost** unsupervised quantum learning algorithm for k-means clustering is a quantum algorithm that uses adiabatic quantum computing to perform clustering on data points. Adiabatic quan-

tum computing is an alternative method of implementing quantum algorithms that involves slowly adjusting the parameters of a quantum system in order to transform a simple initial state into a final state that encodes the result of the computation. The goal of adiabatic quantum computing is to minimize the impact of external noise on the quantum system and preserve the coherence of the quantum state. In the Lloyd, Mohseni and Rebentrost unsupervised quantum learning algorithm for k-means clustering, the quantum system is initialized in a simple ground state and the parameters of the system are slowly adjusted in an adiabatic process. The final ground state of the quantum system encodes the result of the k-means clustering computation. The goal of each clustering step is to create a quantum superposition that encodes the assignment of feature vectors or data points to clusters. This quantum superposition is represented as $|\chi\rangle$, and is defined as a sum of terms that correspond to each data point and the cluster it is assigned to. The authors propose to use adiabatic quantum computing to transform an initial Hamiltonian H_0 into a Hamiltonian H_1 that encodes the distance between a data point and the centroid of the closest cluster. The Hamiltonian H_1 is used to find the optimal assignment of data points to clusters, with the goal of minimizing the distance between each data point and its assigned cluster centroid. In addition to using adiabatic quantum computing for clustering, the authors also mention that this method can be applied to the optimization problem of finding good initial or "seed" centroid vectors. This involves finding the centroid vectors that will result in the optimal assignment of data points to clusters, based on the distance between the data points and the centroids.

2.4 Searching for a quantum neural network model

A challenge for pattern classification with neural networks is the computational cost for the backpropagation algorithm, even when we consider improved training methods such as deep learning.

Elizabeth Behrman is a researcher who has proposed a method for simulating neural networks using interacting quantum dots. A quantum dot is a nanoscale semiconductor structure that can be used to confine and control electrons in a similar way to how a transistor controls the flow of electrons in a conventional computer. Behrman's approach involves using interacting quantum dots to encode and process information in a way that is similar to how neurons and synapses function in the brain. She proposes using the interactions between the quantum dots to simulate the connections between neurons and the transfer of information between them. This approach has the potential to offer significant advantages over classical neural networks, such as the ability to process information in parallel and the ability to perform certain types of computation more efficiently. However, it also comes with some challenges, such as the need to maintain the quantum coherence of the system and the difficulty of scaling up the system to larger sizes.

Fuzzy feed-forward neural networks are a type of machine learning algorithm that are inspired by quantum mechanics and allow for multi-state neurons. These networks are based on the idea of using fuzzy logic, which is a type of mathematical logic that allows for uncertainty and imprecision in the repre-

sentation and processing of information. In fuzzy feed-forward neural networks, the neurons are able to take on multiple states, rather than being limited to just two states like in classical neural networks. This allows for a more flexible and powerful representation of information, and can potentially lead to improved performance on certain tasks. One advantage of using fuzzy logic in neural networks is that it allows for the incorporation of expert knowledge and human decision-making processes into the algorithm. It also allows for the representation of complex and non-linear relationships between the input and output of the network.

Pattern recognition is the process of identifying patterns or regularities in data, and it is an important task in many applications of machine learning. There have been a number of proposals for implementing pattern recognition using adiabatic computing with liquid-state nuclear magnetic resonance (NMR). Adiabatic computing is a method for performing computations using the principles of quantum mechanics. It involves slowly changing the parameters of a quantum system in a way that preserves the system's energy, so that the final state of the system encodes the result of the computation. Adiabatic computing has been proposed as a way to perform certain types of computation more efficiently than is possible using classical computers. Liquid-state NMR is a technique that uses the magnetic properties of certain nuclei in a liquid to perform spectroscopy and other types of measurements. It has been used in a number of applications, including in chemistry and biology to study the structure and function of molecules. There have been a number of proposals for using adiabatic computing with liquid-state NMR to perform pattern recognition tasks. These approaches often involve encoding the patterns to be recognized into the initial state of the NMR system, and then using adiabatic evolution to perform the recognition process. The final state of the system is then measured to obtain the result of the recognition.

2.5 Towards a quantum decision tree

In the design of decision trees, the goal is to select the best decision function for each node in the tree that will split the given dataset into the most organized sub-datasets. One common method for selecting the best decision function is to use a measure of the disorder or randomness of the dataset, such as Shannon's entropy. Shannon's entropy is a measure of the amount of information or uncertainty contained in a dataset. It is defined as the expected value of the amount of information required to specify the value of a random variable. In the context of decision tree design, Shannon's entropy can be used to evaluate the quality of a split by measuring the degree of disorder or randomness in the resulting sub-datasets. The decision function that results in the sub-datasets with the lowest Shannon's entropy is typically selected as the best decision function for the node. This decision function is then used to split the dataset into the most organized sub-datasets, and the process is repeated for each sub-dataset until the tree is complete.

Lu and Brainstein propose a quantum version of the decision tree algorithm for classifying data points using quantum feature states. In this approach, the data points are encoded as quantum states, and the features of the data

points are represented as the states of a quantum system. During the classifying process, the quantum feature states are divided into subsets at each node of the tree by performing a measurement on the quantum states. The subset of quantum states that is selected is then used to determine the next node in the tree. To design the partitioning of the quantum states at each node of the tree, Lu and Brainstein suggest using the von Neumann entropy, which is a measure of the disorder or randomness of a quantum system. The von Neumann entropy can be used to evaluate the quality of the partition by measuring the degree of disorder or randomness in the resulting subsets of quantum states.

2.6 Quantum state classification with Bayesian methods

An important task in quantum information that can be tackled using Bayesian methods is quantum state classification, which involves discriminating between two quantum states produced by an unknown or partly unknown source. This is a classification problem, as the goal is to learn a function that can assign new examples to one of two classes based on their quantum states.

In quantum state classification, the two quantum states are represented by density matrices, which are mathematical objects that describe the statistical properties of a quantum system. A positive operator-valued measurement (POVM) with a binary outcome can be used as a Bayesian classifier to perform the classification.

The basic idea behind using a Bayesian classifier for quantum state classification is to update the probabilities of each class as new information becomes available. This allows the classifier to make more accurate predictions as it gathers more data.

2.7 Hidden quantum Markov models

A hidden quantum Markov model (HQMM) is a type of probabilistic model that represents the evolution of a quantum system over time. It is similar to a classical Markov model, which is a type of probabilistic model that represents the evolution of a system over time, but it takes into account the quantum nature of the system.

In an HQMM, the state of the quantum system at each time step is represented by a density operator, which is a mathematical object that describes the state of a quantum system in terms of the probabilities of different outcomes of measurements on the system. The evolution of the system over time is described by a set of transition operators, which are unitary matrices that describe how the state of the system changes from one time step to the next, just like in a classical Markov model. However, unlike in a classical model, the state of the system is not directly observable in an HQMM. Instead, the state of the system can only be inferred through observations of the system at different times, which can take the form of a set of symbols or measurements.

The observation of the system at a particular time is described by a set of measurement operators, which represent the different ways in which the system can be observed. These measurement operators are determined by the probabilities of observing different symbols or measurements given the current state of the system.