A Comparison of Quantum and Classical Neural Networks in the Supervised Learning

Brian Chen

Qmunity Quantum Fellowship

23 September 2022

ABSTRACT

Quantum computing has been a rising field with promises of bringing innovating improvements to many applications of computer science. One such field, Quantum Machine Learning, involves the implementation of quantum circuits to train neural networks in an attempt to mimic human intelligence. A quantum advantage has yet been reached when comparing the efficiency and performance of state-of-the-art classical models with quantum neural networks . This is largely due to the limitations of qubits available when developing larger models, especially when classical models are able to train billions of parameters, but this poses a question: how do quantum neural networks perform in a comparison featuring similar metrics? A classical model and a quantum model were designed with the same loss function and optimizer, similar structure, and other identical hyperparameters to solve supervised learning problems. The results revealed that under these conditions, a quantum advantage is yet to be seen as the classical model performed with greater accuracy and lower loss values in both regression and classification. However, it is important to note that the common metric chosen are ones that are proven to work well with classical models, but the same cannot be said for quantum neural networks.

INTRODUCTION

The Classical Neural Network

Artificial neural networks are algorithms that are designed to mimic the human brain in solving certain problems. The most basic neural networks feature layers of neurons that fire, like human neurons, based on the input after being processed by their weights and biases and an activation function. The fundamental concept behind the ability of neural networks to solve a wide variety of problems involves the Universal Approximation Theorem. It states that, given enough neurons, each with weights, biases and an activation function, any continuous function can be represented to a reasonably accurate degree [1]. This is the key to the neural network's intelligence, allowing it to learn to compose a function that properly represents the solution to any problem.

One category of machine learning, supervised learning, features training data with corresponding truth labels that the output of the neural network is compared to through a loss function, which calculates the performance of the neural network where a high loss value represents poor performance and a low loss value represents great performance. The loss value is used to calculate the step in which the trainable parameters take through backpropagation, which allows the gradient, representing the loss landscape, to be descended. This is done iteratively until the training process is done or the performance of the neural networks reaches a desired level. However, gradient descent faces challenges when descending loss landscapes that have many local minimums that the model is unable to escape from, which results in stagnation, meaning that the model is unlikely to train more. The training process features many hyperparameters that can affect how the model trains, including the learning rate, which determines how large of a step each iteration will take. Other hyperparameters like batch size (the number of iterations processed at a time) and the number of epochs (the number of iterations over the entire dataset) determine how long the neural network will train for [2].

The Quantum Neural Network

The application of quantum mechanics and quantum computing to machine learning allows for quantum neural networks, otherwise known as quantum variational circuits. There are four approaches to quantum machine learning, depending on the type of data and the type of computing (quantum or classical). The focus of this research is on processing classical data using quantum algorithms.

Quantum neural networks share many similarities with classical neural networks. In fact, some aspects of the training process, such as the calculation of the loss function are done on classical computers. Instead of trainable neurons, the backbone of quantum neural networks is the parameterized quantum variational circuit. Likewise, this circuit contains gates that are tunable by the same process that neurons of classical networks are. In order for these circuits to perform well after training, they need to be able to generate a significant portion of possible states; this measure is known as expressibility. A good parameterized variational circuit should leverage the quantum principle of entanglement [3]. Entanglement between qubits allows the outcome of one to affect another, which after training, the variational circuits are able to leverage this to recognize patterns and effects of those patterns on other qubits that classical models cannot [4].

Encoding Classical Data

Another crucial property of the quantum variation circuit is the ability for classical data, composed of binary bits, to be encoded into quantum data, comprised of qubits that can be in the states, $|0\rangle$, $|1\rangle$, or a combination of the two states. When the qubit is in a combination of the two states, it is said to be in a superposition, which only collapses when the exact state is measured [5]. The measurement of a qubit in superposition functions like probabilities, where the measurement of the same qubit can yield different results. Because of this variant property, quantum algorithms should be run thousands of times in order to produce an accurate distribution of the results. To represent this probabilistic nature of qubits, amplitudes, which consist of a magnitude and a phase, are used to describe the state of qubits. The magnitude is related to the square root of a probabilistic outcome while the phase is used to describe the direction of the state, resulting in a property known as interference, which relates how qubits affect each other when summed [3].

There are many methods of encoding classical bits into qubits, and this is a subfield that may see a lot of innovation research, but one simple method is angle encoding. It converts the data through the following transformation:

$$|x\rangle = \bigotimes_{i=1}^{n} \cos(x_i)\,|0\rangle + \sin(x_i)\,|1\rangle$$

This encodes n features into the rotation angles of n qubits. Compared to other encoding methods, the angle encoding is useful for quantum machine learning as it only utilizes qubits equivalent to the amount of features, allowing it to be an accessible format of encoding that is possible on current quantum computers [3].

METHODS

Datasets

In order to establish a fair comparison of the performance of quantum and classical neural networks on supervised learning problems, controlled variables are crucial. For the regression problem, both models will be trained and tested on the California Housing Price dataset from the 1990 California census. The dataset will be randomly shuffled after each trial and then divided into a training and validation set, but the shuffle will remain the same when both models train and test. The classification problem will utilize the blob generator from sklearn to create blob clusters of training and validation sets. Similarly, each trial will have a new randomly generated and shuffled dataset, but the shuffle and generation will be the same for both models.

Controlled Hyperparameters

In order to establish a fair comparison of the performance of quantum and classical neural networks on supervised learning problems, controlled variables are crucial. Some hyperparameters like the batch size and small changes in learning rate have slight effects on the training of neural networks, but for consistency, a batch size of 32 and a learning rate of 0.03 will be used for all trials, both regression and classification. The classification problems will train for 100 epochs while the regression problem will train for 25 epochs due to the large size of the dataset. The Adam optimizer, which utilizes the previously mentioned gradient descent along with a new concept known as momentum, will be used for all trials. The momentum algorithm is used to accelerate or decelerate the step size based on the exponential moving average [6]. This allows the models to move toward the minimums faster and avoid being trapped in local minima. For this reason, Adam is a popular choice of an optimizer.

Loss Functions

In supervised learning, the choice of loss function is crucial for neural networks to learn. For regression problems, Mean Squared Error (MSE) is the most commonly used loss function where $\hat{Y}_i$ is the predicted value for datapoint X_i and Y_i is the associated truth label:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE, because of the square, punishes large deviations from the truth value while mollifying the loss value if there is a minimal difference [6].

For classification problems, sparse categorical cross entropy is the dominant loss function when classifying into several categories:

$$\text{Loss} = -\sum_{i=1}^{\substack{\text{output}\\\text{size}}} y_i \cdot \log \hat{y}_i$$

Sparse categorical cross entropy calculates the probability of each category and then uses one-hot encoding to compare with the truth label [8].

Control Limitations

Although many aspects of the experiment can be controlled, there are still fundamental differences between classical and quantum models that cannot be controlled. The structure of the two models ultimately cannot be the same, however, some design choices were still made in an attempt for the closest one-to-one comparison. The classical neural network will feature three dense neuron layers, which is the most basic classical structure, while the quantum neural network will utilize a depth of three for each qubit. Since angle encoding was utilized, the number of qubits was determined by the number of features the dataset needed (7 for classification, 8 for regression).

RESULTS

Classification

As seen in figure 1 below, the accuracies of the quantum neural network on the validation tests for classifying blobs is consistently lower than the classical neural network. The average accuracy, across 10 different generated datasets and 5 trials of each, of the quantum neural network is 92.13%, whereas the average accuracy for the classical neural network is 71.10%.

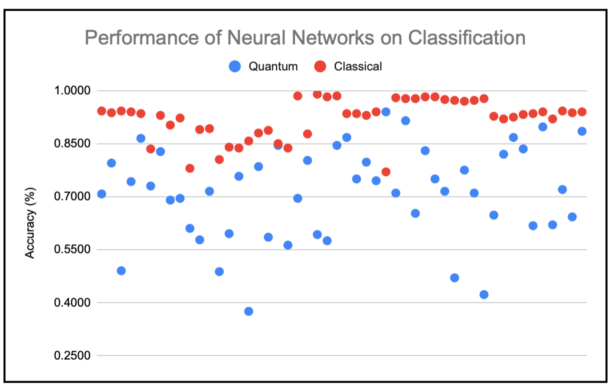

Figure 1. Classification Accuracies of Classical and Quantum Neural Networks

At the same time, the distribution of the quantum neural network varies greatly, with a standard deviation of 13.23%, especially when compared to the classical model's standard deviation of 5.58%. The large spread shows that the quantum model is less consistent, which may be an inherent property of quantum computing due to measurements of superposition or it may be an indicator that the random initializer of weights and biases affects the quantum model more greatly. The performance of the classical model is therefore, more consistent and accurate than the quantum model.

The colormap classifications of both the neural networks may lead to insightful information about how both process or train on classification datasets.

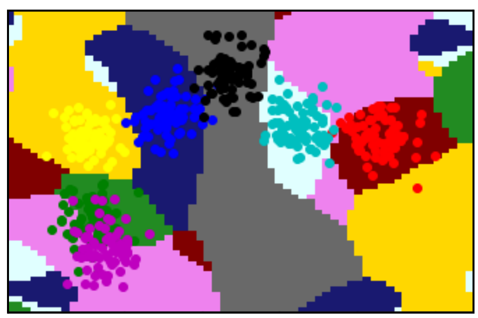

Figure 2. Colormap of Quantum Neural Network on Classifying Clusters

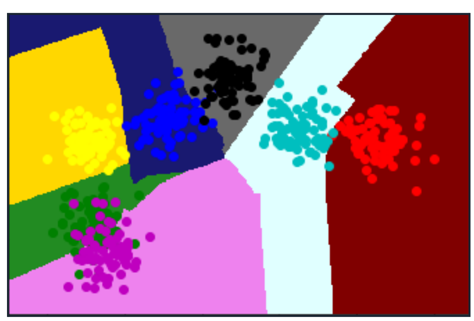

Figure 3. Colormap of Classical Neural Network on Classifying Clusters

Both example colormaps have great performance (QNN: 82.75%, NN: 93.00%), and their accuracy differs slightly, yet their colormaps are drastically different. Figure 2 appears to reveal that the quantum neural network may generalize to zones outside of the dataset more poorly, but also seems to draw less linear boundaries between clusters than the classical model. These patterns are consistent throughout all 50 total trials.

Regression

The performance of the quantum and classical neural networks were both great when addressing the California Housing Regression problem with final average loss values of 0.0297 and 0.0136 respectively.

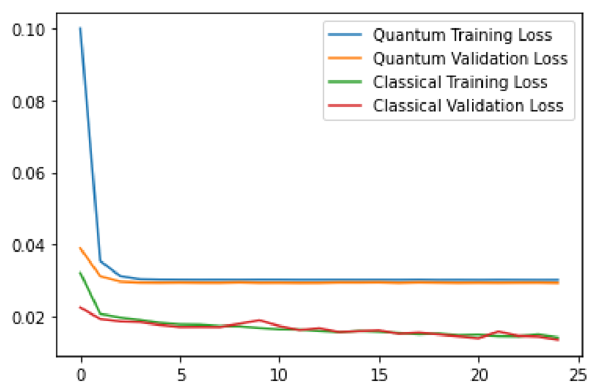

Figure 4. Training History of Neural Networks on California Housing Prices Regression Dataset

With the dataset standardized to be between 0 and 1, the maximum mean squared error is 1, indicating the worst performance whereas a loss of 0 would mean perfect regression. This would reveal that the classical model performs superior to the quantum model, but with loss values so close to 0, both models are comparable in their ability to calculate a function that represents the California housing regression dataset's 8 variables.

The training histories (Figure 4 as an example), reveal that the quantum model consistently reaches its maximum performance rapidly and then stagnates in training whereas the classical model slowly tunes itself ever so slightly.

CONCLUSION

In both categories of supervised learning, regression and classification, the classical model performed better than the quantum model when utilizing the same hyperparameters, similar structure, and the same datasets. If these metrics were the only ones available, the classical neural network would be the superior choice. However, for regression problems, the quantum neural network still performed especially well.

However, it is important to note that these metrics were chosen because it has been proven to work well with classical neural networks. The same cannot be said for quantum neural networks where more optimal choices could exist. Some other choices like the method of encoding and parameterized circuits could also have improvements and did not have equivalents on classical computers. Still, the world of quantum machine learning is promising as more research is done to optimize the structure of the parameterized circuits, hyperparameters, and potentially even loss functions.

REFERENCES

[1] Ye, A. (2021, July 1). *The Universal Approximation Theorem*. Medium. Retrieved September 26, 2022, from https://medium.com/analytics-vidhya/you-dont-understand-neural-networks-until-you-understand-the-universal-approximation-theorem-85b3e7677126

[2] IBM Cloud Education. (n.d.). *What is gradient descent?* IBM. Retrieved September 26, 2022, from https://www.ibm.com/cloud/learn/gradient-descent

[3] *Parameterized quantum circuits*. qiskit.org (n.d.). Retrieved September 26, 2022, from https://learn.qiskit.org/course/machine-learning/parameterized-quantum-circuits

[4] Liu, Y., Li, W.-J., Zhang, X., Lewenstein, M., Su, G., & Ran, S.-J. (1AD, January 1). *Entanglement-based feature extraction by Tensor Network Machine Learning*. Frontiers. Retrieved September 26, 2022, from https://www.frontiersin.org/articles/10.3389/fams.2021.716044/full

[5] Voorhoede, D. (n.d.). *What is a qubit?* Quantum Inspire. Retrieved September 26, 2022, from https://www.quantum-inspire.com/kbase/what-is-a-qubit/

[6] *Intuition of Adam optimizer*. GeeksforGeeks. (2020, October 24). Retrieved September 26, 2022, from https://www.geeksforgeeks.org/intuition-of-adam-optimizer/

[7] *Understanding loss functions in machine Learning*. (n.d.). Retrieved September 26, 2022, from https://www.section.io/engineering-education/understanding-loss-functions-in-machine-learning/#loss-functions-for-regression

[8] *Categorical Cross Entropy Loss Function: Peltarion platform*. Peltarion. (n.d.). Retrieved September 26, 2022, from https://peltarion.com/knowledge-center/modeling-view/build-an-ai-model/loss-functions/categorical-crossentropy