Project: Milestone

Rokin MAHARJAN & Nishan KARKI

July 28, 2022

1 Project Summary

1.1 Overview:

Cluster Analysis is an unsupervised machine learning data analysis technique that is used to find hidden patterns within a given dataset [1]. The popularity of these clustering algorithms is due to their efficiency to find patterns within unlabeled data sets [1]. One such unsupervised machine learning algorithm is K-Means algorithm [2]. K-Means clusters data points based on the pairwise distance between partitions to minimize overall intra-cluster variance [3]. With the adoption of quantum K-Means, we can achieve speed up in distance calculation if we allow the input and output vectors to be quantum states [3]. In this project, we are going to implement the classic K-means Clustering algorithm using quantum computers to cluster unlabeled primate (monkey) data. Our data set has 115 primates' data points with 11 features. We plan to reduce the data dimension using PCA (Principal Component Analyses) and perform K-Means clustering on the resulting data. The distance calculating part of the K-Means algorithm will be implemented using a quantum circuit in qiskit.

1.2 Intellectual Merit:

K-means clustering algorithm is a very time and resource-heavy algorithm and can suffer with huge data sets. With the implementation of K-means in quantum computers, we can drastically speed up the training and prediction time [2]. This project looks forward to adapting classical algorithms to quantum algorithms and dwell upon possibilities of adapting other similar algorithms.

1.3 Broader Impacts of the Proposed Work:

K-Means is a widely popular unsupervised machine learning algorithm. It's uses in the field of biology, big data, chemistry, etc. are well documented. Various techniques have been applied to speed up the run time of similar algorithms. With the adaption of the K-Means algorithm in quantum computing, we might be able to generate faster classical algorithms.

References

[1] Stephen DiAdamo, Corey O'Meara, Giorgio Cortiana, and Juan Bernabé-Moreno. Practical quantum k-means clustering: Performance analysis and applications in energy grid classification, Dec 2021.

[2] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum algorithms for supervised and unsupervised machine learning, Nov 2013.

[3] Peter Wittek. *Quantum machine learning what quantum computing means to data mining.* Elsevier, AP, 2016.