Quantum K-Means

Rokin Maharjan
Nishan Dhoj Karki

Baylor University
08/10/2022

Outline

- Introduction
- Methodology
- Results
- Conclusion and Future Work

Introduction

- K-Means is a widely popular unsupervised machine learning algorithm that assigns data points to a cluster using Euclidean distance measurement.
- K-means can suffer computationally with a large dataset, and most of the computation part happens with Euclidean distance algorithm.
- In this project, we plan to implement Quantum K-Means algorithm with biological (Primates) data and compare the results between classical K-Means and Quantum K-means algorithm.

Methodology - Data Collection

- Hormone data from 110 monkeys
- Monkey's ages ranged from 4 - 10 years
- After feature selection, we have a final dataset with 31 primates with 9 features.
- Most of the data were incomplete to be included in the analysis.

Methodology - Principal Component Analysis

- PCA is an unsupervised linear transformation technique
- It is used for feature extraction and dimensionality reduction
- We reduced our original data dimension: 9 to 2 using PCA

Methodology- K-Means

- K-means is a distance-based clustering algorithm that groups the data into different clusters based on its distance from a center.
- K-means algorithm:
  - Select k - the number of clusters
  - Select k random centroids from the data points
  - Assign each data point to its closest centroid. After this, k clusters will be formed.
  - Calculate the mean of each of the clusters and re-assign the data points to the mean. If any re-assignment is done, repeat this step.
- We used K-means to generate two clusters for LD/BD and HD/VHD

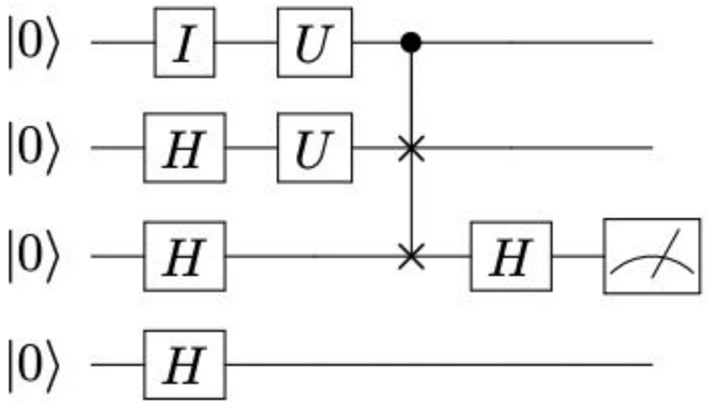
Methodology - Quantum K-means

- K-means is NP-hard with time-complexity of O(t*k*n*d)
- K-means suffers computationally with the increase in the size of data
- With Quantum K-means, we can reduce time complexity to O(N)
- We are implementing the Euclidean distance calculation part using Quantum computing.
- We used Qiskit to implement Quantum K-means

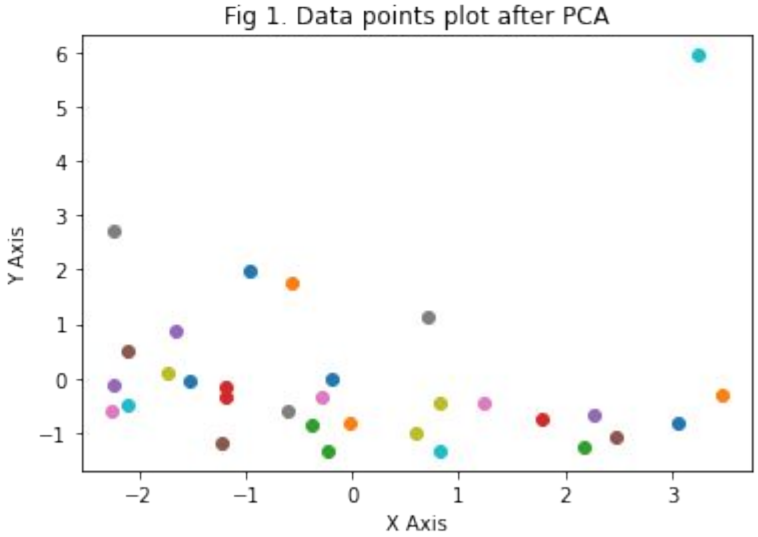
Methodology - Quantum K-means

For our quantum circuit we used 4 quantum registers:

```
qc.h(qr[1])
qc.h(qr[2])
qc.h(qr[3])
qc.u(thetalist[0], philist[0], 0, qr[0])
qc.u(thetalist[i], philist[i], 0, qr[1])
qc.cswap(qr[2], qr[0], qr[1])
qc.h(qr[2])
qc.measure(qr[2], cr[0])
```



Results - After PCA



Results - K-means VS. Quantum K-means

- Accuracy of classical K-means: 67.75%
- Accuracy of quantum K- means: 61.3 %

Fig. 2 Classical K-Means clustering on primates induction data

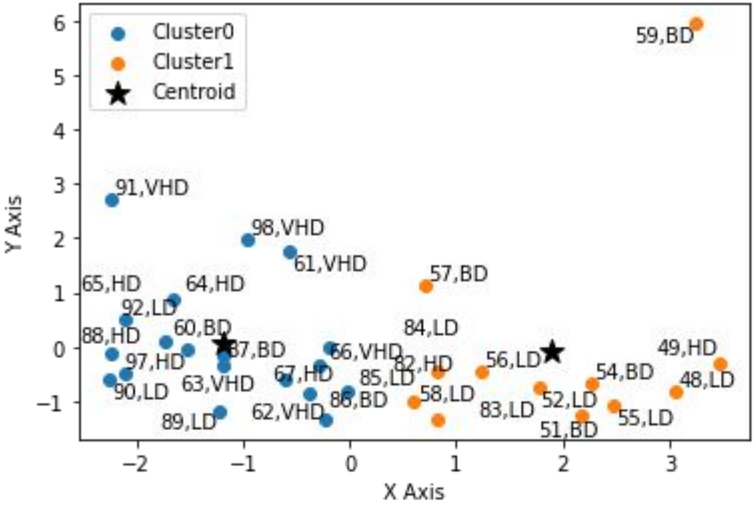
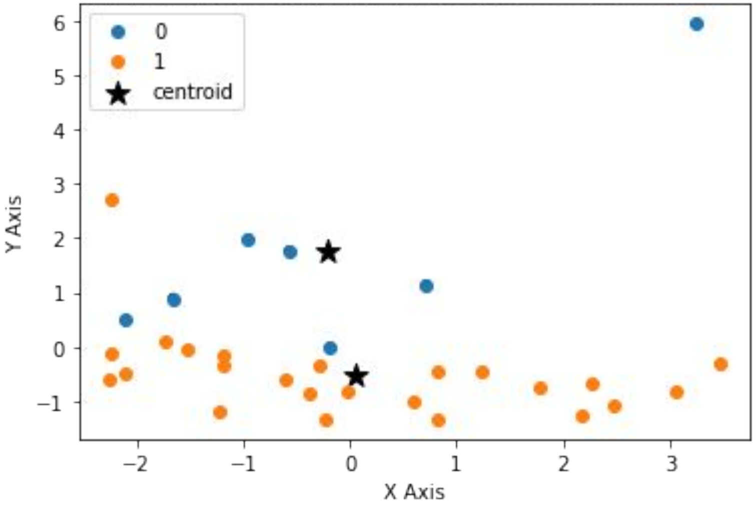


Fig. 3 Quantum K-Means clustering on primates induction data



Conclusion and Future Work

- The accuracy dropped for quantum K-means due to the inefficiency of the circuit
- In the future, we aim to improve the circuit
- We also aim to perform the experiments with larger datasets and more features

References

[1] Erich Baker, Nicole Walter, Alex Salo, Pablo Rivas, Sharon Moore, Steven Gonzales, and Kathleen Grant. Identifying future drinkers: Behavioral analysis of monkeys initiating drinking to intoxication is predictive of future drinking classification. Alcoholism, clinical and experimental research, 41, 01 2017.
[2] Stephen DiAdamo, Corey O'Meara, Giorgio Cortiana, and Juan Bernab é-Moreno. Practical quantum k-means clustering: Performance analysis and applications in energy grid classification, Dec 2021.
[3] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum algorithms for supervised and unsupervised machine learning, Nov 2013.
[4] Sharon Moore, Ami Radunskaya, Elizabeth Zollinger, Kathleen Grant, Steven Gonzales, and Erich Baker. Time for a drink? a mathematical model of non-human primate alcohol consumption. Frontiers in Applied Mathematics and Statistics, 5, 02 2019.
[5] Peter Wittek. Quantum machine learning what quantum computing means to data mining. Elsevier, AP, 2016.


ANY
QUESTIONS?

Thank You!