Quantum Feature Encoding and Dimensionality Reduction
Adequate Module 10
Milestone 1

Muhammad Maaz Arsalan Batla
Ali Muhammad

May 2025

Contents

1 Abstract

In this milestone report, we present the initial stages of our project focused on utilizing quantum computing techniques for data preprocessing in machine learning pipelines.

Specifically, we explore the concepts of quantum data encoding and quantum state tomography as foundational steps towards dimensionality reduction using Quantum PCA.

The objective is to investigate how quantum data encoding methods can effectively transform classical data into quantum states, and how state tomography can be employed to visualize and interpret these quantum states.

The first notebook demonstrates angle encoding and amplitude encoding, converting classical data points into quantum states through rotation gates and amplitude vectors.

The second notebook implements quantum state tomography to reconstruct and visualize these states, offering insights into the fidelity of encoding methods.

Our work aims to establish a foundation for further exploration of quantum preprocessing techniques, setting the stage for the integration of Quantum PCA in the subsequent milestone.

The report highlights key observations, initial findings, and potential challenges encountered during the implementation.

2 Introduction

Classical data preprocessing is a fundamental step in machine learning, regardless of the size and dimensionality of the dataset. Techniques such as feature encoding and dimensionality reduction are employed to simplify data representation while preserving essential patterns and information. However, as data dimensions grow, classical methods may face limitations related to computational complexity and potential data loss.

Quantum computing introduces new paradigms for data processing by encoding classical data into quantum states[1], allowing for more compact data representations[2] and potentially faster computations. The ability to encode data as quantum states not only reduces dimensionality[3][4] but also leverages quantum properties like superposition and entanglement, which can enable more efficient data processing.

In this project, we focus on two key quantum preprocessing techniques: angle and amplitude encoding, and quantum state tomography. In the first notebook, we implement angle encoding and amplitude encoding using Qiskit, demonstrating how classical data points can be represented as quantum states through rotation gates and amplitude vectors. We explore the pros and cons of each method, considering factors such as data capacity, computational complexity, and visualization.

The second notebook introduces quantum state tomography, a method used to reconstruct quantum states based on measurement data. This technique allows us to visualize encoded quantum states and assess the fidelity of encoding methods. By reconstructing the quantum states and comparing them with their ideal counterparts, we gain valuable insights into the accuracy and effectiveness of the encoding techniques.

This milestone report summarizes the work completed in these two notebooks, providing theoretical explanations, implementation details, and initial observations. It also establishes the groundwork for the next phase of the project, which involves Quantum PCA and its application in quantum-enhanced ML pipelines.

3 Notebook 1: Angle and Amplitude Encoding

3.1 Objective

The primary objective of this notebook is to introduce and implement two fundamental methods for encoding classical data into quantum states: angle encoding and amplitude encoding[5]. These methods are crucial for converting classical data into quantum representations that can be processed by quantum algorithms. By exploring both encoding techniques, we aim to compare their effectiveness in representing data and visualizing the encoded states on the Bloch sphere.

3.2 Theory and Concepts

In classical machine learning terminology, a **data point** or **sample** refers to an individual item in a dataset—such as a single image, document, or measurement. Each data point is described by a set of **features**, for example: pixel intensity (x_1), texture (x_2), or edge count (x_3). Thus, a sample is typically represented as a feature vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]$.

Quantum encoding techniques aim to map such a feature vector into a quantum state $|\psi\rangle$. We explore two encoding strategies:

Angle Encoding: In angle encoding, each individual feature x_i is mapped to the rotation angle of a quantum gate (commonly R_y or R_z) applied to a corresponding qubit. This means that the number of qubits required equals the number of features in the data point:

$$|\psi\rangle = R_y(x_1)|0\rangle \otimes R_y(x_2)|0\rangle \otimes \cdots \otimes R_y(x_n)|0\rangle \tag{1}$$

This approach is easy to implement and interpret, especially for low-dimensional data. However, its scalability is limited due to the one-to-one mapping between features and qubits.

Amplitude Encoding: Amplitude encoding maps the entire feature vector into the amplitudes of a quantum state. For a normalized feature vector $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ satisfying $\sum |x_i|^2 = 1$, the encoding becomes:

$$|\psi\rangle = \sum_{i=0}^{N-1} x_i |i\rangle \tag{2}$$

Here, $|i\rangle$ represents computational basis states of $n = \log_2 N$ qubits. This means amplitude encoding can represent 2^n features using only n qubits, making it highly efficient for large feature vectors. However, this method is more complex in terms of state preparation and typically requires entangled operations and precise normalization.

3.3 Implementation

We used Qiskit[6] to implement both encoding methods. For angle encoding, each feature was mapped to a rotation using the R_y gate on a separate qubit. For amplitude encoding, the normalized feature vector was directly encoded into a quantum register using the `initialize()` function. We simulated the resulting states using AerSimulator and visualized them with Bloch spheres[7] to observe the quantum state configuration.

3.4 Results and Analysis

The Bloch sphere visualizations showed how the two methods differ in their data representations. Angle encoding resulted in separate rotations of individual qubits, providing a straightforward but resource-intensive visualization. Amplitude encoding compressed the entire feature vector into a single state over fewer qubits, but required precise normalization and was more sensitive to numerical errors.

In summary, angle encoding is suitable for small, low-dimensional datasets due to its simplicity, while amplitude encoding is more powerful for larger feature sets where quantum resource efficiency becomes a concern.

4 Notebook 2: Quantum State Tomography

4.1 Objective

The objective of this notebook is to implement quantum state tomography to reconstruct and visualize quantum states encoded using angle and amplitude encoding methods. This technique allows for a detailed analysis of the fidelity of encoded states and helps in verifying the accuracy of the encoding process.

4.2 Theory and Concepts

Quantum state tomography[8] is a method used to reconstruct the density matrix of a quantum state based on measurement data. The density matrix ρ is a mathematical representation of a quantum state, capturing all the information required to describe the state:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{3}$$

In this notebook, we utilize Qiskit's tomography modules to reconstruct the density matrix and compare it with the ideal state to assess fidelity[9]. The fidelity F between two states ρ and σ is defined as:

$$F(\rho, \sigma) = \left(\mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right)^2 \tag{4}$$

A fidelity value closer to 1 indicates a higher similarity between the encoded and reconstructed states.

4.3 Implementation

The implementation involves generating a quantum state using angle and amplitude encoding, performing tomography to reconstruct the state, and calculating fidelity to quantify the accuracy of the reconstruction. The results are visualized using Bloch spheres and density matrices, providing a comprehensive understanding of the encoded states.

4.4 Results and Analysis

The results demonstrate the reconstructed states and their fidelity values. A higher fidelity indicates that the encoded states are accurately reconstructed, validating the effectiveness of the encoding methods. Observations are provided to highlight the strengths and limitations of the implemented techniques.

5 Challenges and Issues

Throughout the development of the first two notebooks, several challenges and issues were encountered. One significant challenge was ensuring the correct implementation of amplitude encoding, which required precise normalization of data vectors. Improper normalization resulted in errors during state initialization, affecting the statevector generation and subsequent visualization.

Another notable issue arose during the implementation of the tomography reconstruction in the second notebook. The reconstruction process involved generating multiple measurement bases and combining the results to construct the density matrix. Ensuring accurate fidelity calculations required careful alignment of the reconstructed state with the original encoded state. In some cases, small discrepancies in measurement data led to lower fidelity values, prompting us to refine the circuit design and adjust measurement settings.

Additionally, integrating Qiskit's AerSimulator and resolving compatibility issues with specific Qiskit functions posed initial implementation hurdles. Adjustments to function calls, simulator backends, and transpilation processes were necessary to ensure accurate statevector generation and Bloch sphere visualization.

Despite these challenges, the successful implementation of angle encoding, amplitude encoding, and state tomography established a strong foundation for the next phase of the project, where Quantum PCA will be introduced to further analyze encoded data states.

6 Conclusion and Next Steps

The completion of the first milestone marks significant progress in our exploration of quantum data preprocessing techniques. The implementation of angle and amplitude encoding provided a comprehensive understanding of how classical data can be effectively mapped to quantum states, with visualizations illustrating the resulting statevectors on the Bloch sphere. Quantum state tomography further enabled the reconstruction of encoded states, allowing for a detailed analysis of fidelity and the accuracy of the encoding methods.

Looking ahead, the next phase of the project will focus on integrating Quantum Principal Component Analysis (QPCA) to analyze the encoded data states and reduce dimensionality. Additionally, we plan to incorporate machine learning[10] pipelines that leverage the encoded quantum states as input features, enabling a comparative analysis of classical and quantum preprocessing methods in ML workflows.

The next milestone will involve further research on the theoretical foundations of QPCA, implementation of the algorithm in Qiskit, and a comprehensive evaluation of the preprocessing pipeline. By systematically extending the quantum feature encoding techniques explored thus far, we aim to establish a robust framework for quantum-enhanced data preprocessing and dimensionality reduction.

7 References

References

[1] QuantumAlgorithms.org, "Quantum algorithms for data analysis," 2024, accessed: 2025-05-16. [Online]. Available: https://quantumalgorithms.org/chap-dimensionality-reduction.html

[2] Q. Hu and L. Chen, "Guided quantum compression for high dimensional data classification," *Journal of Quantum Computing*, vol. 7, no. 2, pp. 150–162, 2024, accessed: 2025-05-16. [Online]. Available: https://iopscience.iop.org/article/10.1088/2632-2153/ad5fdd/pdf

[3] Fiveable, "Quantum dimensionality reduction methods - fiveable study guide," 2024, accessed: 2025-05-16. [Online]. Available: https://library.fiveable.me/quantum-machine-learning/unit-13/quantum-dimensionality-reduction-methods/study-guide/9q2SeJmDolaD8AwW

[4] J. Liu and W. Zhang, "Quantum dimensionality reduction by linear discriminant analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 614, 2023, accessed: 2025-05-16. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0378437123001097

[5] L. Wang and F. Zhang, "Quantum data encoding techniques," *arXiv preprint arXiv:2404.03256*, 2024, comprehensive overview of encoding methods. [Online]. Available: https://arxiv.org/abs/2404.03256

[6] Q. Community, *A Practical Guide to Qiskit: Quantum Computing for Developers.* IBM Research, 2024, comprehensive guide for implementing quantum algorithms. [Online]. Available: https://qiskit.org/documentation/

[7] Q. D. Team, "Bloch sphere visualizations in qiskit," 2024, guide on visualizing quantum states. [Online]. Available: https://qiskit.org/documentation/visualizations/bloch_sphere.html

[8] D. Smith and A. Lee, "An introduction to quantum state tomography," *Quantum Information Processing*, vol. 22, no. 3, pp. 45–60, 2023, explains state reconstruction techniques. [Online]. Available: https://link.springer.com/article/10.1007/s11128-023-3584-1

[9] M. Johnson and A. Patel, "Fidelity measures for quantum state comparisons," *Journal of Quantum Computing*, vol. 7, no. 1, pp. 112–130, 2023, analysis of fidelity calculations and applications. [Online]. Available: https://doi.org/10.1016/j.jqc.2023.01.003

[10] Spinquanta, "Quantum machine learning explained: From theory to use," 2024, accessed: 2025-05-16. [Online]. Available: https://www.spinquanta.com/news-detail/quantum-machine-learning-explained-from-theory-to-use