



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

Анализ данных калибровки кубитов: использование гауссовой смеси и линейной регрессии для классификации и визуализации

Студент ИУ5-32М
(Группа)

(Подпись, дата) А.В. Воронцова
(И.О.Фамилия)

Руководитель

(Подпись, дата) Ю.Е. Гапанюк
(И.О.Фамилия)

2024 г.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ
Заведующий кафедрой ИУ5
(Индекс)
В.И. Терехов
(И.О.Фамилия)
«__» _____ 2024 г.

ЗАДАНИЕ
на выполнение научно-исследовательской работы

по теме Анализ данных калибровки кубитов: использование гауссовой смеси и линейной регрессии для классификации и визуализации

Студент группы ИУ5-32М

Воронцова Алина Владимировна
(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)
ИССЛЕДОВАТЕЛЬСКАЯ

Источник тематики (кафедра, предприятие, НИР) КАФЕДРА

График выполнения НИР: 25% к ____ нед., 50% к ____ нед., 75% к ____ нед., 100% к ____ нед.

Техническое задание Разработать алгоритм классификации на основе гауссовой смеси и линейной регрессии для двух состояний кубита, оценить эргономику решения

Оформление научно-исследовательской работы:

Расчетно-пояснительная записка на 27 листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

Дата выдачи задания «__» _____ 2024 г.

Руководитель НИР

Ю.Е. Гапанюк
(Подпись, дата) (И.О.Фамилия)

Студент

А.В. Воронцова
(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ.....	6
1.1. Современные подходы к калибровке кубитов.....	6
1.2. Методы сбора данных: усреднение и single-shots.....	6
1.3. Применение гауссовой смеси и линейной регрессии в анализе данных.....	7
2. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ.....	10
2.1. Подготовка данных: структура и форматы.....	10
2.2. Построение и обучение моделей гауссовой смеси.....	11
2.3. Построение и обучение модели линейной регрессии.....	16
2.4. Предсказание модели GMM.....	17
2.5. Визуализация результатов.....	20
2.6. Анализ и сравнение.....	22
ЗАКЛЮЧЕНИЕ.....	26
СПИСОК ЛИТЕРАТУРЫ.....	27

ВВЕДЕНИЕ

Квантовые вычисления представляют собой одно из самых перспективных направлений современной науки и технологий[1], открывая новые возможности для решения задач, которые недоступны классическим вычислительным системам. Основным элементом квантового компьютера является кубит[2] — физическая система, обладающая уникальной способностью находиться в состоянии суперпозиции и выполнять вычислительные операции на основе квантовой механики. Однако для корректной работы кубита требуется предварительная калибровка, в ходе которой определяются параметры управляющих сигналов, обеспечивающие точное выполнение вычислений. Одним из ключевых параметров является амплитуда сигналов, которая напрямую влияет на надежность работы кубита.

Процесс калибровки [4] кубитов сопровождается анализом данных, полученных в ходе многократных измерений. Эти данные могут собираться двумя основными способами: с усреднением результатов измерений или с использованием единичных снимков (single shots). Первый способ, связанный с усреднением, обеспечивает общую картину поведения кубита, однако может сглаживать важные детали, которые могут оказаться критически значимыми для точной настройки. Второй способ, основанный на методе single shots, позволяет фиксировать отдельные измерения, предоставляя детальную информацию о состоянии кубита в каждом измерении. Именно такой подход наиболее точно отражает поведение системы и выявляет скрытые метаморфозы, которые не очевидны при использовании усреднений.

Актуальность данного исследования определяется необходимостью разработки методов анализа данных, собранных методом single shots, для повышения точности и надежности процесса калибровки кубитов. Такой подход позволяет находить оптимальные параметры управления квантовой системой, что критически важно для успешного выполнения вычислений, например: подобрать корректную амплитуду для работы.

Целью настоящей работы является разработка алгоритмов анализа данных, полученных методом single-shots, преимущественно, с использованием моделей гауссовой смеси (GMM) [5], а также визуализация результатов анализа для упрощения интерпретации. Для достижения поставленной цели требуется решить следующие задачи:

1. Реализовать алгоритмы обработки данных, основанные на методе single shots, с применением GMM.
2. Разработать модели распределения данных, описывающие состояния кубита в основном и возбужденном состояниях.
3. Использовать логарифмы правдоподобия для классификации данных и выбора оптимального кейса калибровки.
4. Создать инструменты визуализации, включая построение эллипсов ковариации и отображение расстояний между состояниями, для наглядного представления результатов анализа.
5. Провести анализ данных для определения оптимальной амплитуды управляющего сигнала.

В рамках исследования также проводится сравнение двух методов для классификации оптимальной амплитуды калибровки: моделей гауссовой смеси (GMM) и линейной регрессии с признаками. GMM позволяет гибко моделировать многокомпонентные распределения данных и выявлять скрытые закономерности, в то время как линейная регрессия подходит для линейных зависимостей, но ограничена в учете нелинейных отношений, что может снижать точность при анализе сложных данных.

Практическая значимость работы особенно проявляется в контексте эргономики процесса калибровки квантовых систем. В условиях лабораторий, таких как ФМНС, выбор параметров амплитуды зачастую осуществляется вручную, "на глаз", что не позволяет учесть тонкие особенности в данных и может снижать точность вычислений.

1. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1.1. Современные подходы к калибровке кубитов

Калибровка кубитов — это фундаментальная задача в квантовых вычислениях, от которой напрямую зависит точность выполнения операций и надежность квантовой системы. Калибровка включает настройку управляющих параметров, таких как амплитуда, частота и длительность импульсов, взаимодействующих с кубитом. Эти параметры определяют эффективность квантовых операций, включая выполнение элементарных логических операций и реализацию схем квантового алгоритма. [3]

Современные подходы к калибровке базируются на экспериментах с измерением отклика кубита на различные управляющие сигналы. В научной литературе описаны два основных направления. Первое предполагает использование усредненных данных, где многократные измерения объединяются для уменьшения шумов и выявления общего поведения системы. Второе направление связано с использованием данных отдельных измерений, или *single-shots*, что позволяет фиксировать индивидуальные состояния кубита и анализировать их поведение на микроуровне.

Эффективность калибровки во многом определяется алгоритмами обработки данных. Традиционно калибровка производилась вручную, однако растущее количество параметров и сложность квантовых систем требуют автоматизации процессов. Одной из ключевых задач остается разработка методов, которые учитывают скрытые закономерности в данных, недоступные для анализа "на глаз".

1.2. Методы сбора данных: усреднение и *single-shots*

Методы сбора данных при калибровке кубитов играют решающую роль в обеспечении точности измерений. Метод усреднения используется для получения общей картины поведения кубита, снижая влияние флуктуаций и шума. В этом подходе каждый измерительный сигнал регистрируется

многократно, после чего результаты объединяются, вычисляя их среднее значение. Такой подход позволяет уменьшить влияние случайных шумов и обеспечивает более стабильное отображение поведения системы.

Однако усреднение скрывает тонкие метаморфозы, которые могут быть ключевыми для точной настройки. Для выявления таких изменений применяют метод *single-shots*, позволяющий регистрировать индивидуальные измерения. Этот метод считается более чувствительным к микроскопическим изменениям в состоянии кубита и лучше отражает реальные особенности его поведения. Метод *single-shots* представляет собой анализ каждого индивидуального измерения отклика кубита без усреднения. Он фиксирует каждый "снимок" состояния кубита, позволяя детально изучить его поведение. *Single-shots* требуют высокой точности измерительных приборов, поскольку каждое измерение несет информацию о текущем состоянии кубита.

Стоит подчеркнуть, что использование данных *single-shots* особенно актуально в условиях работы с многокубитными системами, где требуются высокая точность при калибровке. Однако данный метод требует значительных вычислительных и временных ресурсов.

1.3. Применение гауссовой смеси и линейной регрессии в анализе данных

Модели гауссовой смеси (GMM, Gaussian Mixture Models) представляют собой мощный метод для анализа данных, позволяющий описывать сложные распределения на основе их представления в виде смеси нормальных распределений. В рамках данной задачи GMM применяется не только для анализа данных *single-shots*, но и для обучения модели классифицировать файлы данных, снятых при разных амплитудах управляющего сигнала, с целью определения оптимальной калибровки кубита.

В отличие от простых методов классификации, GMM не требует четкого разбиения данных на кластеры. Вместо этого модель обучается на данных, описывая их вероятностную структуру, и использует параметры нормальных распределений для оценки, насколько хорошо данные каждого

файла соответствуют различным моделям. Для каждого файла данных вычисляется логарифм правдоподобия, который отражает, насколько вероятно, что эти данные относятся к определенной амплитуде калибровки. GMM характеризуется следующими ключевыми параметрами:

1. Среднее значение (μ): представляет центр распределения данных.
2. Ковариационная матрица (Σ): описывает форму, масштаб и наклон распределения.
3. Вес (π): отражает вклад каждой компоненты смеси в общую модель.

Используя эти параметры, GMM строит описание для каждого файла данных, чтобы затем сравнить их и определить, к какому варианту калибровки файл относится.

В задаче калибровки кубитов важно определить, какая амплитуда управляющего сигнала приводит к наиболее точным переходам между основным и возбужденным состояниями. Для этого снимаются данные при разных амплитудах, которые затем обрабатываются и анализируются.

Одной из полезных функций GMM является возможность визуализации данных. Для каждой модели можно построить эллипсы ковариации, отражающие распределение данных для соответствующей амплитуды. Эти визуализации помогают понять, насколько данные файла близки к определенной модели, и интерпретировать результаты анализа. В рамках данной задачи очень удобно проверять перекрытие облаков точек двух состояний или отслеживать выбросы, так как GMM позволяет указать при построении эллипсов область в сигмах (метод трех сигм).

В отличие от методов, использующих гауссовы смеси, линейная регрессия является более простым и понятным инструментом для анализа данных. Она моделирует зависимость между входными и выходными переменными через линейную функцию. В контексте задачи калибровки кубитов линейная регрессия может применяться для анализа отклонений между координатами в основном и возбужденном состояниях кубита, а также для предсказания значений одной координаты на основе другой.

Линейная регрессия в этой задаче может быть использована для следующих целей:

1. Изучение взаимосвязи между состояниями кубита: Например, можно исследовать зависимость между координатами в возбужденном состоянии (e_x и e_y) и управляющим сигналом.
2. Оценка отклонений: Линейная регрессия позволяет вычислить разницу между фактическими значениями и теми, которые предсказаны моделью, что может помочь в выявлении аномальных данных и выбросов.

Однако линейная регрессия имеет свои ограничения, так как она предполагает линейные зависимости между переменными. Это становится проблемой, если данные обладают более сложной структурой или если они следуют многокомпонентным распределениям, которые невозможно адекватно описать с помощью линейной модели.

Преимущество линейной регрессии заключается в ее простоте и прозрачности. В отличие от гауссовых смесей, которые используют вероятностные распределения, линейная регрессия непосредственно показывает зависимость одной переменной от другой, что делает результаты анализа более понятными.

Таким образом, линейная регрессия и гауссова смесь представляют собой два разных подхода в анализе данных. GMM более гибка и мощная, так как может учитывать сложные структуры, такие как многокомпонентные распределения, и позволяет визуализировать результаты через эллипсы ковариации. Линейная регрессия, в свою очередь, простая в интерпретации, но ограничена в описании более сложных зависимостей. Оба метода могут быть использованы для анализа данных, снятых при разных амплитудах управляющего сигнала, с целью нахождения оптимальной калибровки кубита, и выбор между ними зависит от сложности данных и необходимой точности анализа.

2. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

2.1. Подготовка данных: структура и форматы

Для проведения эксперимента и анализа данных были использованы файлы, содержащие информацию о измерениях квантовых состояний кубитов при различных амплитудах управляющего сигнала. Эти данные сняты с помощью метода single-shot, что позволяет зафиксировать квантовое состояние кубита за одно измерение, не влияя на его дальнейшую эволюцию. Каждый файл включает два типа данных: для основного состояния (ground state) и возбужденного состояния (excited state).

Пример структуры данных файла:

1. `g_x`: координаты основного состояния кубита по оси X.
2. `g_y`: координаты основного состояния кубита по оси Y.
3. `e_x`: координаты возбужденного состояния кубита по оси X.
4. `e_y`: координаты возбужденного состояния кубита по оси Y.

Каждое измерение — это массив чисел, представляющих значения координат для состояния кубита в конкретный момент времени. Данные для основного и возбужденного состояний кубита хранятся отдельно в массивы `g_x`, `g_y`, `e_x` и `e_y`, каждый из которых содержит одно измерение в двухмерном пространстве. Рассмотрим структуру данных на примере конкретного файла `Probe_amp_sweep_0.11V_processed.pkl`.

Примерные значения данных для файла:

```
{'g_x': array([-35.00180427, -84.22168629, -62.97458871, ..., -59.23083723,
              -85.67628799, -81.02446413]),
 'g_y': array([ 7.37402813, -31.12490475, -16.00507632, ..., 25.72947561,
              -15.70759662, -1.33497551]),
 'e_x': array([-39.86753243, 19.30344023, 31.77336416, ..., 18.25203534,
              25.66550564, -46.01811667]),
 'e_y': array([-28.11621179, 36.3592404 , 45.88647551, ..., 38.30668124,
              22.82095697, 7.54293185])}
```

Говоря о формате, данные представлены в виде файлов с расширением .pkl (формат сериализации Python), что является стандартом для сохранения объектов Python, включая массивы и структуры данных. Эти файлы содержат сериализованные данные и могут быть загружены с помощью библиотеки pickle. Пример загрузки данных из файла можно посмотреть на рис.1.

```
import pickle

with open('/content/drive/MyDrive/ФМН/Массивы_сингл_шотов/Case 1/Probe_amp_sweep_0.11V_processed.pkl', 'rb') as f:
    data = pickle.load(f)
data

{'g_x': array([-35.00180427, -84.22168629, -62.97458871, ..., -59.23083723,
              -85.67628799, -81.02446413]),
 'g_y': array([ 7.37402813, -31.12490475, -16.00507632, ..., 25.72947561,
              -15.70759662, -1.33497551]),
 'e_x': array([-39.86753243, 19.30344023, 31.77336416, ..., 18.25203534,
              25.66550564, -46.01811667]),
 'e_y': array([-28.11621179, 36.3592404 , 45.88647551, ..., 38.30668124,
              22.82095697, 7.54293185])}
```

Рис.1 - Пример загрузки данных из файла

Загрузив данные, можно приступить к их обработке, анализу и применению различных методов для классификации и предсказания амплитуды, которая оптимально соответствует калибровке кубита. Однако не стоит забывать про предварительную обработку и подготовку данных. Для дальнейшего анализа и использования моделей машинного обучения данные из массивов g_x, g_y, e_x, e_y могут быть преобразованы в единую структуру. Пример объединения данных в двумерные массивы расположен на рис.2.

```
import numpy as np

# Объединяем данные для основного и возбужденного состояния
ground_data = np.column_stack((g_x, g_y))
excited_data = np.column_stack((e_x, e_y))
```

Рис.2 - Пример объединения данных в двумерные массивы

2.2. Построение и обучение моделей гауссовой смеси

Для начала опишем процесс построения и обучения моделей гауссовой

смеси (GMM) [5] для калибровки кубитов, а также их сохранение для дальнейшего использования. В контексте работы с данными, полученными методом single-shot, такие модели помогают более точно классифицировать различные состояния кубитов и определить оптимальные параметры калибровки.

Для обучения моделей GMM используется информация, полученная из нескольких файлов, каждый из которых содержит данные о координатах кубита в двумерном пространстве для двух состояний: основного (ground state) и возбужденного (excited state). Эти данные были собраны из различных опытов, и каждый файл представляет собой результат одного измерения, сохраненный в формате .pkl (Python pickle). Основные данные для обучения модели:

1. g_x , g_y : координаты кубита в основном состоянии.
2. e_x , e_y : координаты кубита в возбужденном состоянии.

Для каждого из кейсов (например, Case 1, Case 2, Case 3) данные из всех файлов собираются в один общий массив, что позволяет создать достаточно обширную выборку для обучения моделей. Важно отметить, что разметка Case 1-Case 3 несет в себе смысл выдуманна не просто так.

Case 1 представляет собой ситуацию, когда амплитуда считывания была правильно подобрана. Это идеальный случай, при котором данные отображают четкое разделение между состояниями кубита — основным и возбужденным. При адекватной амплитуде измерения создают данные, которые находятся в явном и четком распределении в двумерном пространстве (например, координаты g_x , g_y для основного состояния и e_x , e_y для возбужденного). В этом случае состояния кубита хорошо различимы друг от друга, и мы получаем хорошие результаты при их классификации с помощью алгоритмов, таких как модель гауссовой смеси (GMM). Пример таких данных изображен на рисунке 3.

Case 2 демонстрирует ситуацию, когда амплитуда считывания была завышена. Это ведет к тому, что состояния кубита начинают сильно «разлетаться» друг от друга. На графике данные будут демонстрировать более

широкие и не столь четкие области для каждого из состояний, что обусловлено слишком большой амплитудой. В этом случае распределения для основного и возбужденного состояний становятся более растянутыми, и между ними возникает значительное расстояние. Кроме того, при завышенной амплитуде может возникать еще одна проблема — выбросы. Эти выбросы будут появляться как аномальные данные, отклоняющиеся от основной массы точек, создавая дополнительные «хвосты» в распределении. Пример таких данных изображен на рисунке 4.

Case 3 характеризуется тем, что амплитуда считывания была слишком низкой. В этом случае кубиты не могут полностью «разделиться» в пространстве состояний, и данные для основного и возбужденного состояний оказываются сильно «склееными». Это приводит к тому, что два состояния практически неразличимы друг от друга, и данные для них накладываются. На графике это будет выглядеть как одно большое скопление точек, в котором нет явных границ между состояниями. Пример таких данных изображен на рисунке 5.

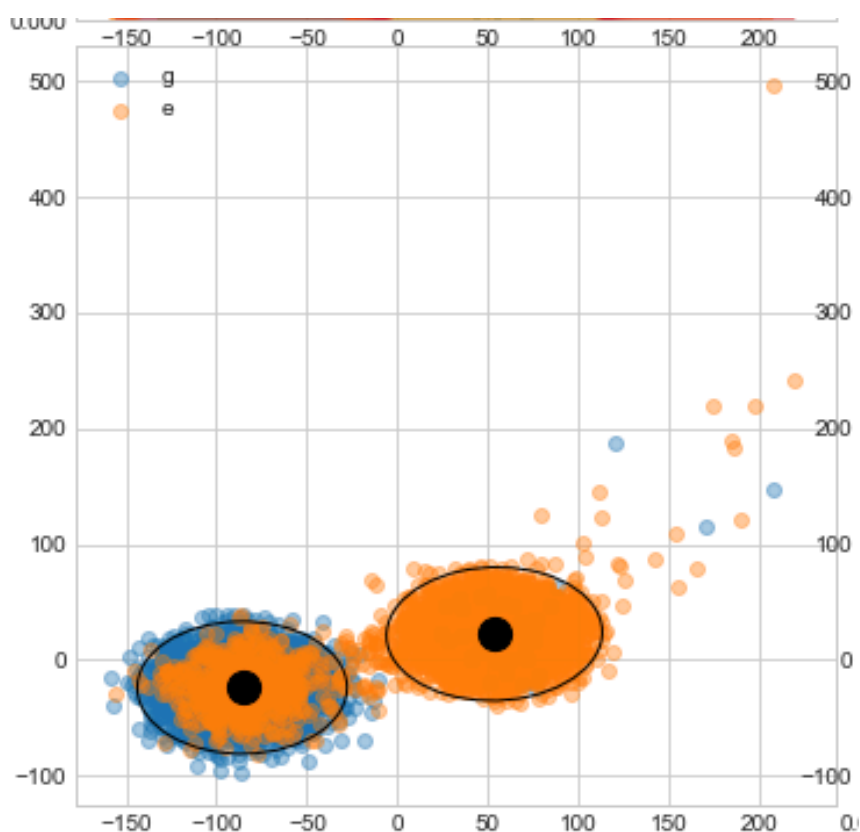


Рис.3 - Распределение облаков точек для Case 1

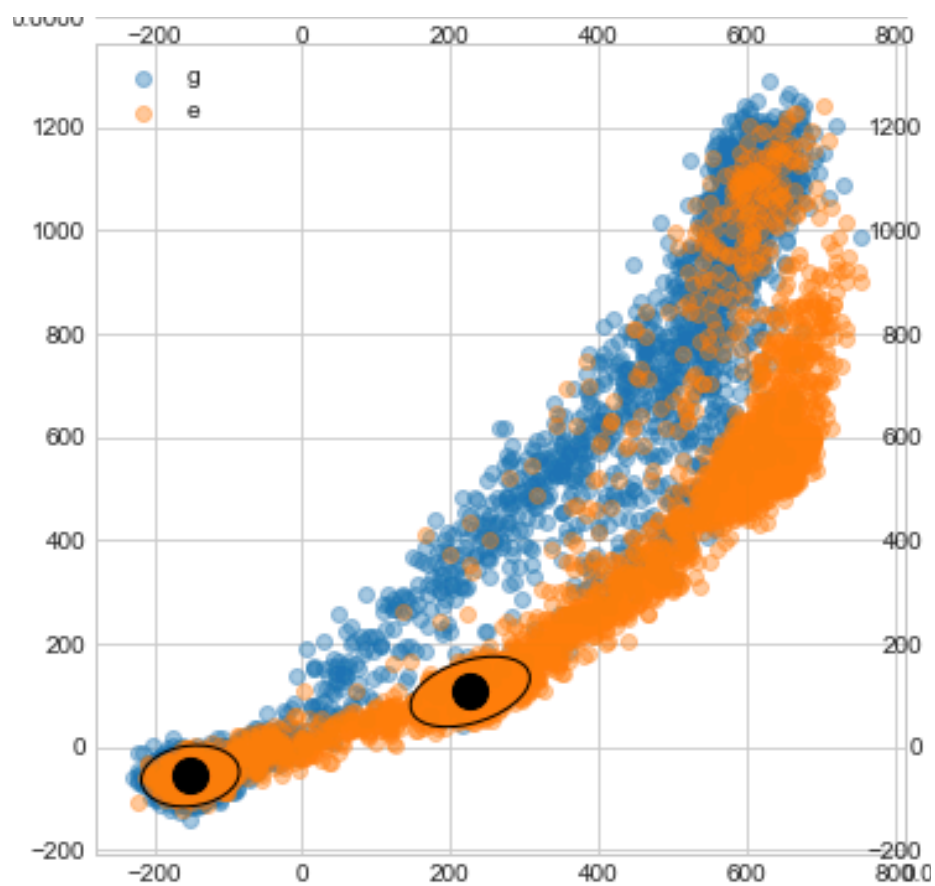


Рис.4 - Распределение облаков точек для Case 2

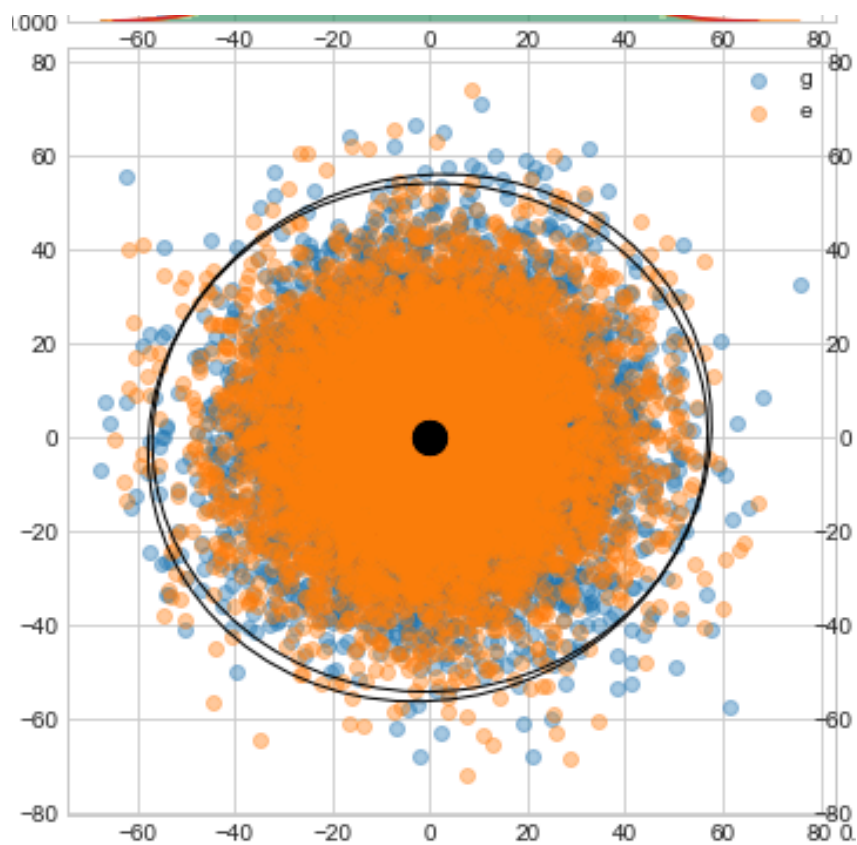


Рис.5 - Распределение облаков точек для Case 3

Модель гауссовой смеси (GMM) предполагает, что данные являются результатом смешивания нескольких нормальных распределений. В нашем случае мы используем одну компоненту для каждого состояния кубита, то есть мы предполагаем, что данные для основного и возбужденного состояний могут быть описаны одним нормальным распределением с определенными параметрами. Обучение модели можно увидеть на рисунке 6.

```
# Обучаем модели GMM
gmm_ground = GaussianMixture(n_components=1, covariance_type='full').fit(ground_data)
gmm_excited = GaussianMixture(n_components=1, covariance_type='full').fit(excited_data)
```

Рис.6 - Пример обучения модели

Здесь `n_components=1` указывает на то, что для каждого состояния используется одно нормальное распределение. Параметр `covariance_type='full'` позволяет модели учитывать полную ковариацию между осями X и Y для каждого состояния.

После того как модели обучены, их необходимо сохранить, чтобы использовать в дальнейшем для предсказаний. Модели сохраняются в формате `.pkl` с использованием библиотеки `joblib`.

Таким образом, для каждого кейса сохраняются две модели: одна для основного состояния и одна для возбужденного. Итог обучения моделей отображен на рисунке 7.

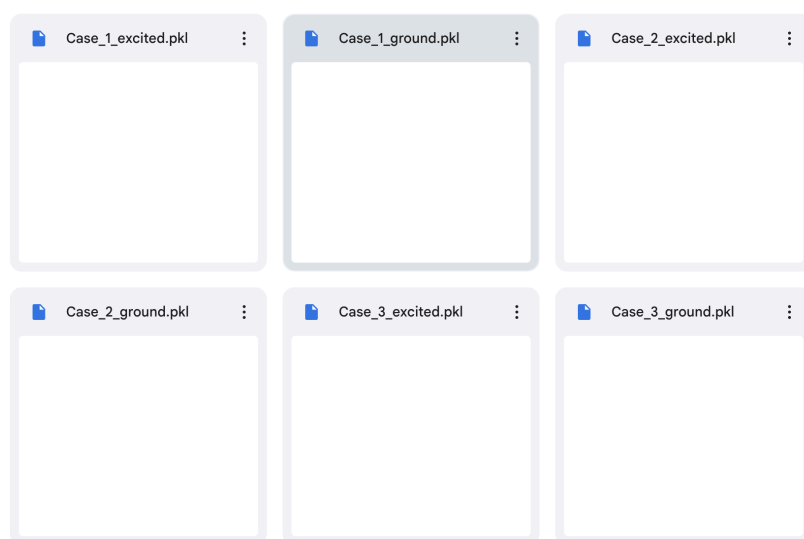


Рис.7 - Обученные модели GMM для Case 1 - Case 3

2.3. Построение и обучение модели линейной регрессии

В этой главе рассмотрим процесс построения и обучения модели линейной регрессии для анализа данных, полученных в ходе калибровки кубитов. Линейная регрессия используется для выявления зависимостей между характеристиками, такими как координаты в основном и возбужденном состояниях кубита, и для предсказания отклонений между этими состояниями.

Первым шагом является загрузка данных, которые хранятся в подкаталогах определенной структуры. В этом примере используется метод `load_data_from_folders`, который загружает данные из файлов `.pkl` в указанных папках. Для каждого файла извлекаются координаты в основном (g_x, g_y) и возбужденном (e_x, e_y) состояниях, после чего рассчитываются следующие признаки:

1. Средние значения для координат в обоих состояниях:
 - a. $\text{mean_g_x}, \text{mean_g_y}$ — средние значения для координат в основном состоянии.
 - b. $\text{mean_e_x}, \text{mean_e_y}$ — средние значения для координат в возбужденном состоянии.
2. Стандартные отклонения:
 - a. $\text{std_g_x}, \text{std_g_y}$ — стандартные отклонения для координат в основном состоянии.
 - b. $\text{std_e_x}, \text{std_e_y}$ — стандартные отклонения для координат в возбужденном состоянии.
3. Максимальное расстояние между координатами двух состояний, которое вычисляется как:

$$\text{max_distance} = \max(\sqrt{(e_x - g_x)^2 + (e_y - g_y)^2}) \quad (1)$$

4. Отклонение, рассчитанное с использованием линейной регрессии:

$$\text{deviation} = \frac{1}{N} \sum_{i=1}^N |e_{y,i} - \overline{e_{y,i}}| \quad (2)$$

, где $\hat{e}_{y,i}$ — предсказанное значение координаты e_y по линейной модели.

Эти признаки используются для классификации данных, определяя, к какому кейсу калибровки (например, "Case 1", "Case 2", "Case 3") относится каждый файл данных.

После подготовки данных следующим шагом является разделение их на обучающую и тестовую выборки с использованием функции `train_test_split` из библиотеки `sklearn`. В данном примере 80% данных используется для обучения, а 20% — для тестирования модели. Для классификации данных используется модель случайного леса (Random Forest), которая обучается на признаках и метках классов, полученных на предыдущем шаге.

Процесс обучения модели с использованием случайного леса заключается в следующем:

1. Модель случайного леса создается с помощью класса `RandomForestClassifier`, где устанавливается фиксированное значение `random_state` для воспроизводимости результатов.
2. Модель обучается на обучающих данных с использованием метода `fit`.
3. После обучения модель используется для предсказания меток на тестовых данных с помощью метода `predict`.
4. Результаты предсказания оцениваются с помощью метрик точности (`accuracy_score`) и отчета о классификации (`classification_report`), который включает такие показатели, как точность, полнота и F-меры для каждого класса.

2.4. Предсказание модели GMM

Для предсказания необходимо использовать ранее обученные модели GMM, которые были сохранены для каждого из кейсов. Эти модели сохранялись в файлы с помощью библиотеки `joblib` после этапа обучения. Для каждого кейса были обучены две модели:

1. Модель для основного состояния (ground).

2. Модель для возбужденного состояния (excited).

В коде загружаются модели для всех трёх кейсов: Case_1, Case_2, и Case_3. Пути к моделям сохраняются в переменных model_ground_path и model_excited_path и передаются в функцию joblib.load(), которая загружает сохраненные модели.

Для предсказания необходимо подать на вход модель новый тестовый файл, содержащий измеренные данные для основного и возбужденного состояний. В коде тестовый файл загружается с помощью функции pd.read_pickle(). Данные в тестовом файле должны содержать четыре ключевых массива:

1. g_x и g_y — координаты для основного состояния.
2. e_x и e_y — координаты для возбужденного состояния.

Если тестовый файл не содержит всех необходимых ключей, код генерирует ошибку и останавливает выполнение. Это предотвращает дальнейшие шаги с некорректными данными.

После загрузки данных и моделей для каждого кейса вычисляется сумма логарифмов правдоподобия для двух моделей (для основного и возбужденного состояний). Логарифм правдоподобия оценивает, насколько вероятно, что данные были сгенерированы текущей моделью. Чем выше логарифм правдоподобия, тем более вероятно, что данные принадлежат данной модели.

Функция calculate_log_likelihood() принимает на вход две модели GMM (для основного и возбужденного состояний) и данные тестового файла.

Она вычисляет логарифм правдоподобия для обеих моделей с использованием метода .score() модели GMM, который возвращает логарифм правдоподобия для переданных данных. Логарифм правдоподобия для смеси гауссовых распределений можно выразить следующей формулой:

$$\log p(x|\theta) = \sum_{i=1}^N \log \left(\sum_{k=1}^K \pi_k \cdot N(x_i | \mu_k, \Sigma_k) \right) \quad (3)$$

, где:

x_i - i -й элемент данных,

θ — параметры модели,

K — количество компонент в смеси (например, количество нормальных распределений в модели GMM),

N — количество наблюдений (данных), для которых оценивается правдоподобие.

Для каждого состояния (основного и возбужденного) вычисляется свой логарифм правдоподобия, после чего они суммируются для каждого кейса.

После того как для каждого кейса вычислены логарифмы правдоподобия, результаты сравниваются. Модель с наибольшим значением логарифма правдоподобия считается наиболее подходящей для тестового файла. В коде это реализовано с помощью функции `np.argmax()`, которая находит индекс наибольшего значения в списке логарифмов правдоподобия. Результат предсказания выводится в консоль в виде сообщения, которое можно увидеть на рис.8.

```
/usr/local/lib/python3.10/dist-packages/sklearn
https://scikit-learn.org/stable/model\_persiste
warnings.warn(
Case_1: Log-likelihood = -18.80
/usr/local/lib/python3.10/dist-packages/sklearn
https://scikit-learn.org/stable/model\_persiste
warnings.warn(
Case_2: Log-likelihood = -25.05
/usr/local/lib/python3.10/dist-packages/sklearn
https://scikit-learn.org/stable/model\_persiste
warnings.warn(
Case_3: Log-likelihood = -20.94

--- Prediction Result ---
The test file most likely belongs to: Case_1
```

Рис.8 - Вывод результата

2.5. Визуализация результатов

Для оценки характеристик данных и результатов обучения моделей был использован метод визуализации, включающий построение графиков распределений для каждого состояния кубита (основного и возбужденного). В процессе анализа каждого файла с данными, выполненного с использованием метода гауссовой смеси (GMM), были построены диаграммы рассеяния с наложенными эллипсами, которые отражают области вероятности для основного и возбужденного состояний. Эти визуализации помогают наглядно понять, как различаются распределения данных и насколько сильно они перекрываются для разных состояний.

Каждая диаграмма рассеяния сопровождается эллипсами, соответствующими основному и возбужденному состояниям, а также информацией о расстоянии между центрами распределений. На графиках строятся эллипсы, которые соответствуют области вероятности, заданной для каждого состояния. Этими эллипсами отображаются области, внутри которых находится 99.7% всех точек для каждого из состояний. Это достигается за счет использования трех сигм (3σ) от центра распределения для каждой из осей эллипса.

Почему используются именно три сигмы? В статистике 99.7% всех точек в нормально распределенной выборке находятся в пределах трех стандартных отклонений от среднего значения (т.е. в пределах трех сигм). Таким образом, эллипс, построенный на основе ковариационной матрицы с учетом трех сигм, отображает зону, которая включает почти все точки, принадлежащие этому распределению.

Примеры графиков для Case 1- Case 3 приведены на рис.9-11.

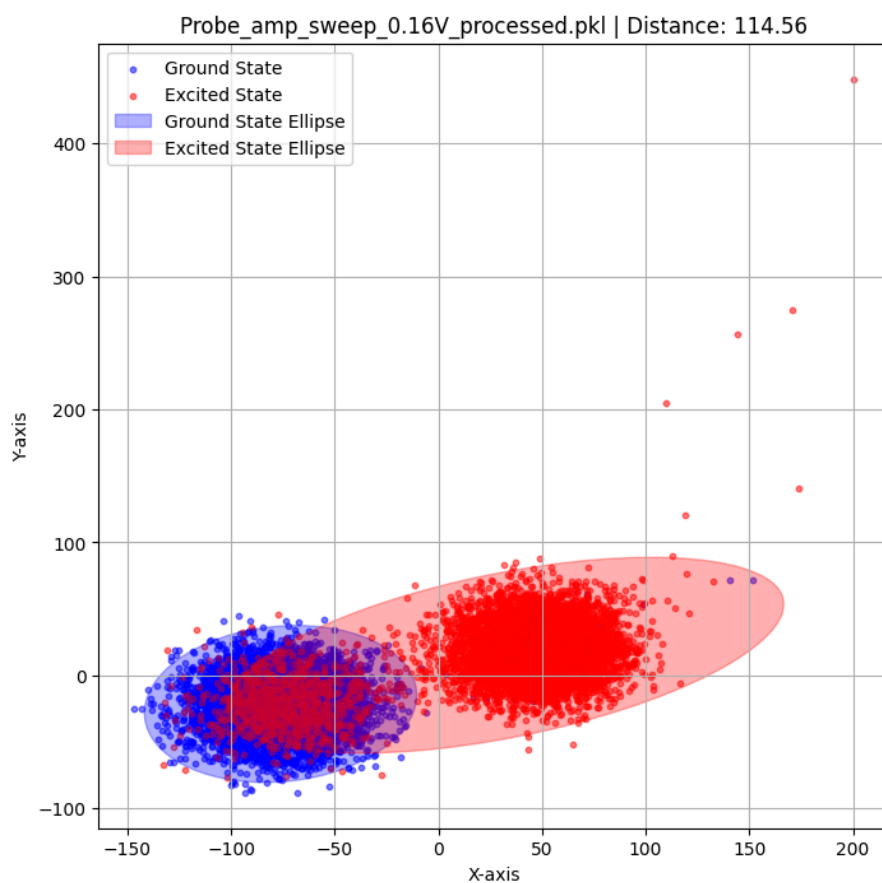


Рис.9 - Результат работы GMM для Case 1

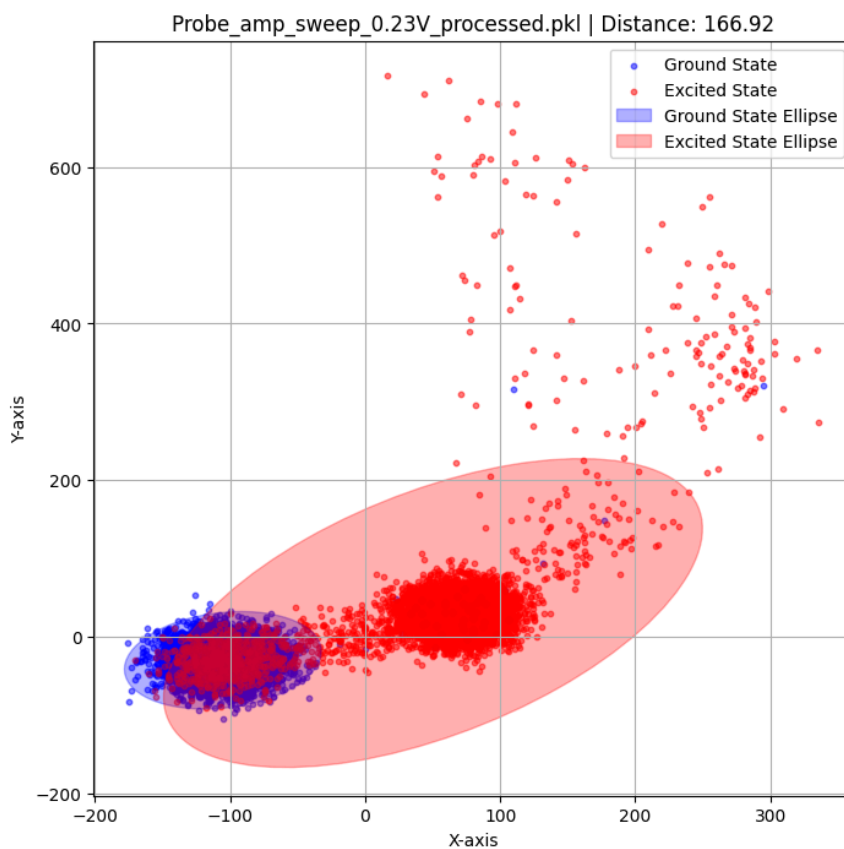


Рис.10 - Результат работы GMM для Case 2

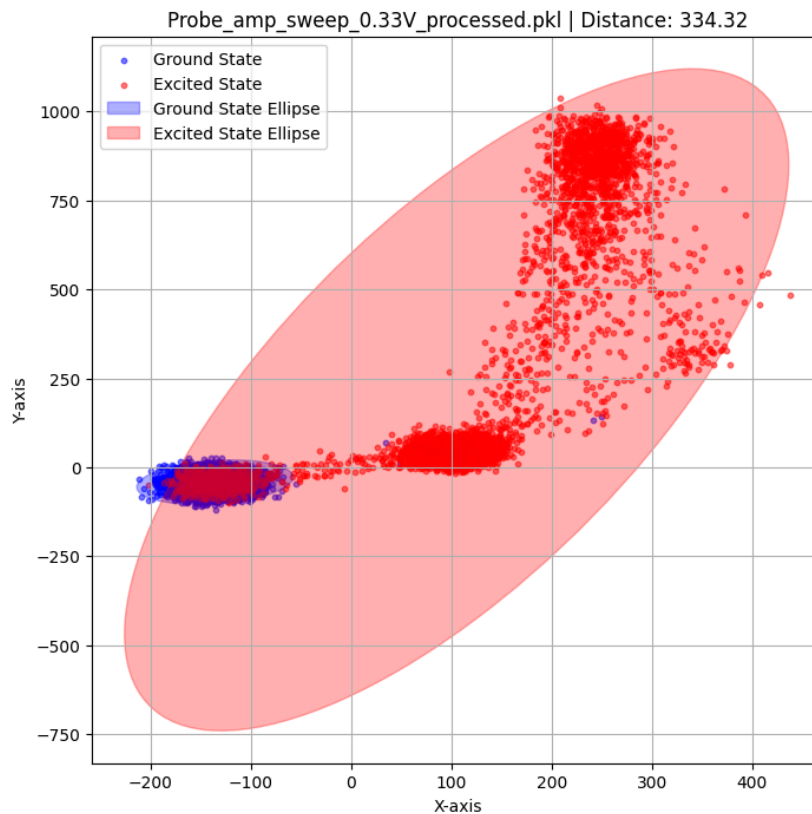


Рис.11 - Результат работы GMM для Case 3

2.6. Анализ и сравнение

В ходе проведенного исследования были проанализированы данные, полученные при разных амплитудах, варьирующихся от 0.01 до 0.89.

На основе этих данных было выявлено, что оптимальной амплитудой является 0,22, так как при значении 0,23 наблюдается резкий скачок в распределении данных. Примеры графиков для амплитуд 0,22 и 0,23 приведены на рисунках 12 и 13. Однако, на данный момент GMM ошибается при классификации оптимальной амплитуды, поскольку она находит корректную амплитуду 0,22, но также ошибочно относит амплитуду 0,26 к числу допустимых значений.

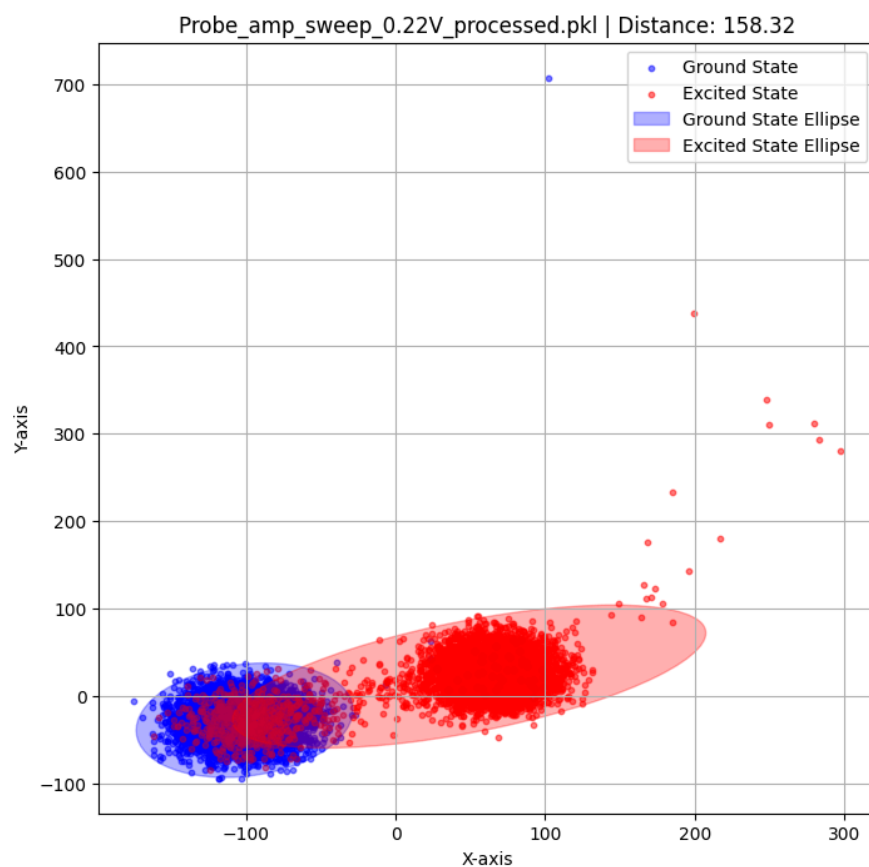


Рис.12 - Оптимальная амплитуда для работы - 0.22

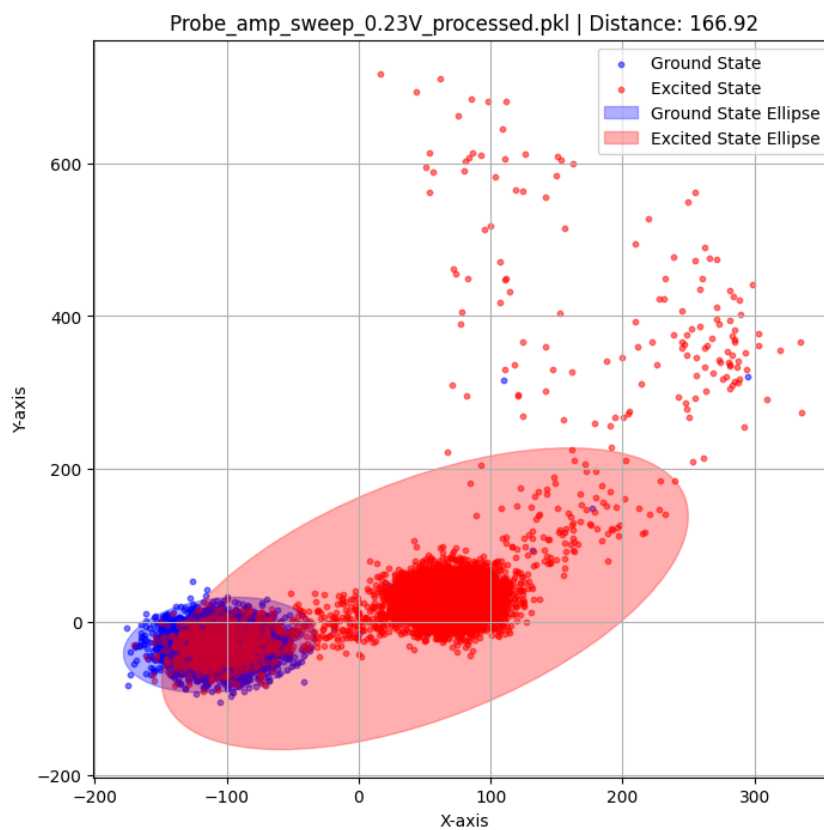


Рис.13 - Результат для пограничной амплитуды 0.23

С точки зрения визуального анализа, амплитуду 0,26 нельзя отнести к оптимальному диапазону, так как она явно завышена и не соответствует поведению данных в рамках данного кейса. Предполагается, что эта ошибка связана с недостатком данных для качественного обучения модели. GMM обучается лишь на 6 файлах хороших данных, что недостаточно для корректного построения надежной модели. При этом модель также обучается на 20 файлах, которые содержат менее точные или даже шумные данные. Это приводит к тому, что GMM может неправильно интерпретировать распределения, особенно при попытке идентифицировать оптимальную амплитуду.

В качестве альтернативы для сравнения результатов с моделью гауссовой смеси (GMM) была написана модель линейной регрессии с признаками. Для классификации оптимальной амплитуды калибровки использовались следующие признаки:

1. Среднее значение координат в состоянии управляющего сигнала (g_x и g_y) — эти характеристики помогают понять общие тенденции в распределении значений сигнала в базовом состоянии.
2. Среднее значение координат в возбужденном состоянии (e_x и e_y) — показывают изменения в поведении кубита при возбуждении.
3. Стандартное отклонение для g_x , g_y , e_x и e_y — эти признаки информируют о степени изменчивости данных.
4. Максимальное расстояние между точками в двух состояниях (между g_x , g_y и e_x , e_y) — позволяет измерить различия между состояниями.
5. Отклонение по линейной регрессии — рассчитывается как среднее абсолютное отклонение между фактическими и предсказанными значениями, что показывает, насколько хорошо линейная модель описывает поведение системы.

Эти признаки были выбраны, поскольку они содержат информацию о различиях между состояниями кубита и дают полезную информацию для различия между кейсами с разными амплитудами управляющего сигнала.

Статистические характеристики (средние и стандартные отклонения) показывают общие тенденции в данных, а максимальное расстояние и отклонения по линейной регрессии помогают выявить различия в поведении кубита в разных состояниях.

Линейная регрессия и гауссова смесь оба являются статистическими методами, но они различаются по подходу:

1. Линейная регрессия ищет линейные зависимости между входными и выходными данными. Она хороша для простых зависимостей, когда можно предположить прямолинейную связь между переменными.
2. Гауссова смесь (GMM) — это более гибкий метод, который моделирует данные как смесь нескольких нормальных распределений. Он может выявлять более сложные структуры и многокомпонентные распределения в данных, что позволяет лучше моделировать поведение системы с несколькими возможными состояниями.

Обе модели, несмотря на различия в подходах, показали одинаковые результаты при классификации оптимальной амплитуды калибровки. Оба метода привели к выводу, что оптимальная амплитуда составляет 0,22. Пример результата работы линейной регрессии приведен на рисунке 14.

```
file_path = "/content/drive/MyDrive/ФМН/Массивы_сингл_шотов/Case 1/Probe_amp_sweep_0.22V_proces
model_path = "/content/drive/MyDrive/ФМН/Массивы_сингл_шотов/trained_model.pkl"
predict_case(model_path, file_path)
```



```
Model loaded from /content/drive/MyDrive/ФМН/Массивы_сингл_шотов/trained_model.pkl
Predicted case: Case 1
'Case 1'
```

Рис.14 - Результат работы модели линейной регрессии

ЗАКЛЮЧЕНИЕ

В ходе эксперимента для амплитуд в диапазоне от 0.01 до 0.89 была определена оптимальная амплитуда для работы с кубитами — 0.22. При этой амплитуде данные показывают наилучшее разделение состояний кубитов (основного и возбужденного), без значительных выбросов или перекрытий. Однако, начиная с амплитуды 0.23, наблюдается резкий рост разброса данных, с появлением «хвостов» на графиках распределений, что указывает на ухудшение качества разделения состояний.

Линейная регрессия и модель гауссовой смеси (GMM) продемонстрировали одинаковые результаты, подтверждая, что амплитуда 0.22 является оптимальной для минимизации отклонений и улучшения разделения состояний кубитов. Для амплитуд выше 0.22 оба метода показали значительные отклонения между предсказанными и фактическими значениями, что свидетельствует о ухудшении качества модели.

Однако, несмотря на сходные результаты, GMM требует более тщательной настройки и обучения. В частности, она иногда добавляет ложные амплитуды в кейс 1, что может указывать на недостаток данных для обучения модели. Возможно, линейная регрессия в данном контексте лучше захватывает зависимости по признакам, предоставляя более точные результаты при

На данный момент лаборатория использует амплитуду 0.19 для работы, несмотря на то, что она не является оптимальной с точки зрения разделения состояний кубитов. Выбор этой амплитуды был сделан на основании анализа усредненных данных, с учетом соотношения сигнал/шум.

СПИСОК ЛИТЕРАТУРЫ

1. Душкин Р. В. Квантовые вычисления и функциональное программирование. — 2014. — 318 с., ил.
2. Соловьев В. М. Квантовые компьютеры и квантовые алгоритмы. Часть 1. Квантовые компьютеры // Математика. — 2014. — № X. — С. 123–145.
3. Как работают квантовые компьютеры: что нужно знать и как они изменят будущее // Habr. — 2020. — [Электронный ресурс]. — URL: <https://habr.com/ru/companies/ruvds/articles/488342/> (дата обращения: 25.12.2024).
4. Superconducting Qubits — Qubit Spectroscopy // Zhinst. — [Электронный ресурс]. — URL: https://docs.zhinst.com/labone_q_user_manual/applications_library/how-to-guides/sources/01_superconducting_qubits/01_workflows/03_qubit_spectroscopy.html (дата обращения: 25.12.2024).
5. Gaussian Mixture Models // Scikit-learn. — [Электронный ресурс]. — URL: <https://scikit-learn.ru/stable/modules/mixture.html> (дата обращения: 25.12.2024).