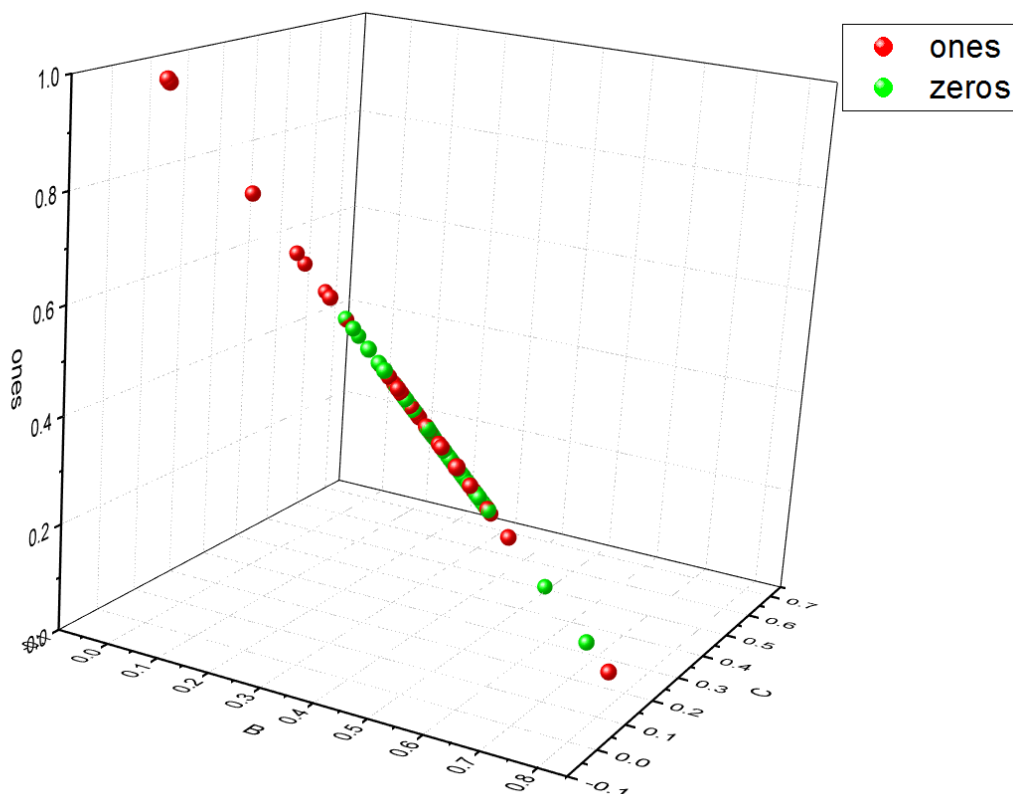
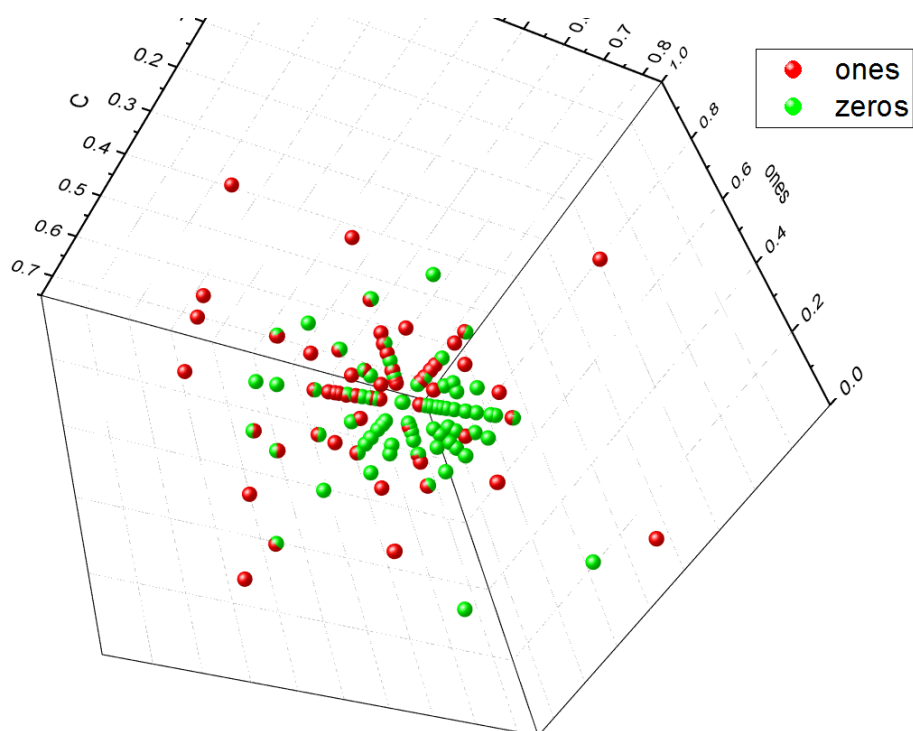


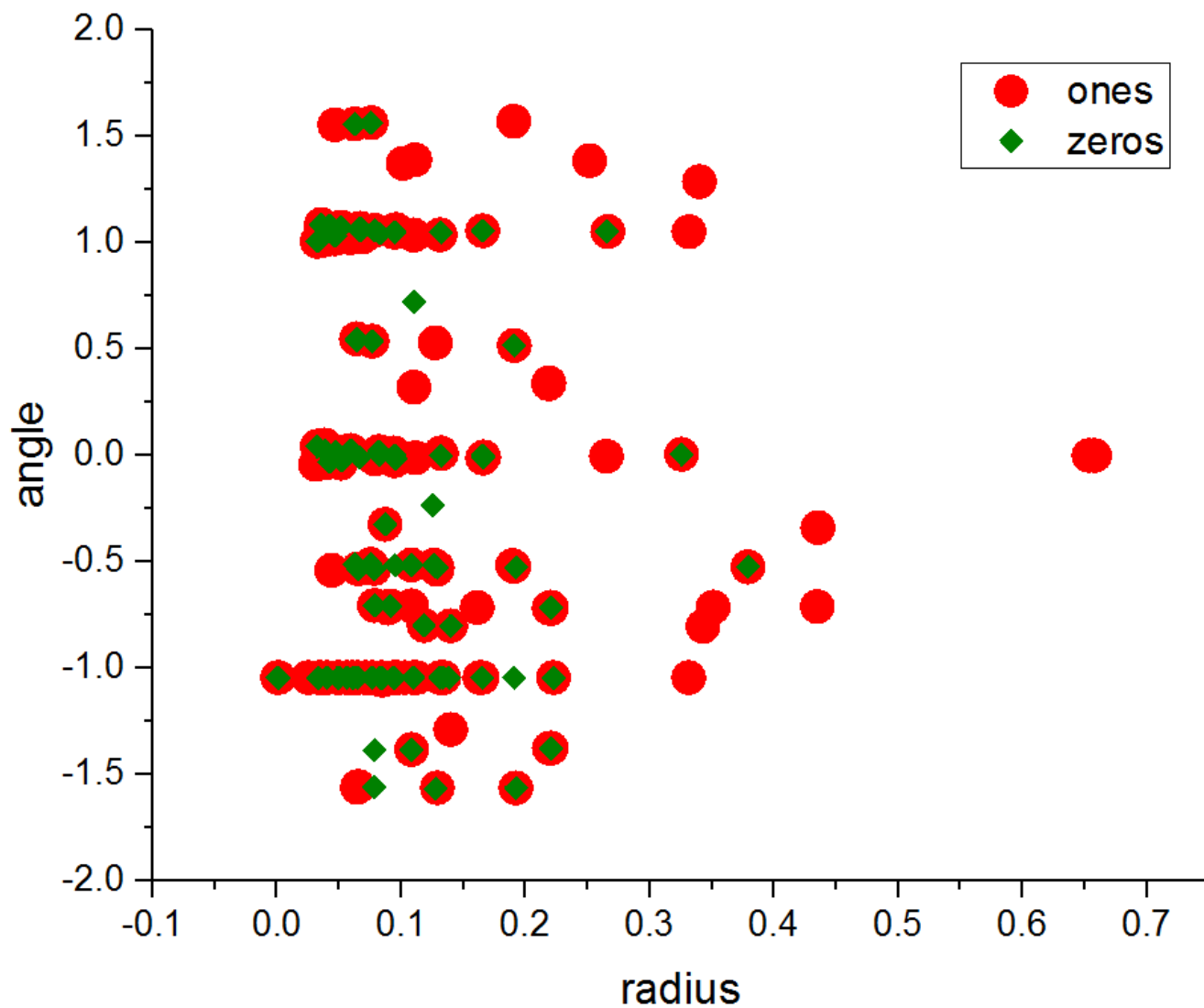
Были выбраны случайные пять файлов с данными из архива.  
Нулевые строки (их много особенно в первой половине файла) были удалены.  
Было отмечено, что данные лежат в плоскости (или очень близко к ней).



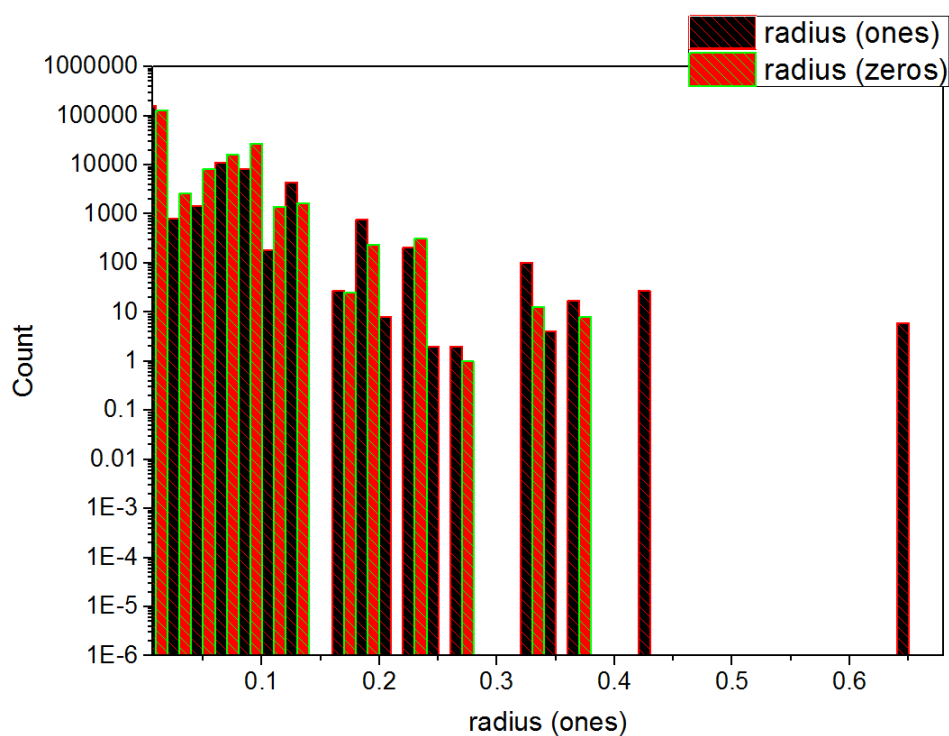
Также было отмечено, что данные имеют радиальную симметрию.

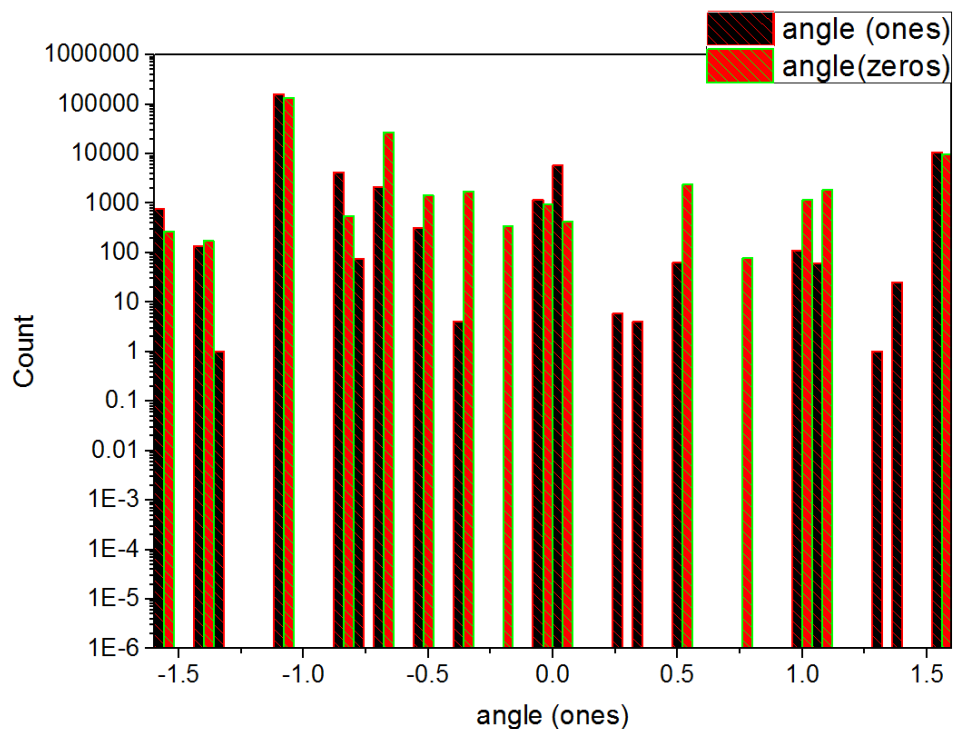


Поэтому была найдена плоскость, из условия минимума суммы квадратов расстояния от точек до плоскости. На эту плоскость были спроецированы точки, найден их центр "звездочки". Он был сдвинут в начало координат, был сделан переход в полярную систему координат.



Анализ зависимостей углов и радиусов для точек разных классов не позволили сделать вывод о наличии очевидных кластеров для точек одного класса. Хотя видно, что существуют направления, вдоль которых есть точки только одного класса. Но доля их мала ( $< 1\%$ ).





Была сделана попытка применить методы машинного обучения для классификации и кластеризации модифицированных данных. Были добавлены новые признаки, полученные нелинейными преобразованиями исходных признаков и их комбинаций. Применялись различные методы классификации. Лучший результат был получен на деревьях решений. Лучший результат не превысил 80%.

1486416673294.csv

Decision tree: Accuracy: 0.5919 (+/- 0.2834)

1486430023690.csv

Decision tree: Accuracy: 0.7958 (+/- 0.2036)

1486435241183.csv

Decision tree: Accuracy: 0.6667 (+/- 0.4273)

1486444008135.csv

Decision tree: Accuracy: 0.7722 (+/- 0.2149)

1486448918687.csv

Decision tree: Accuracy: 0.6667 (+/- 0.4273)