

实验设计

实验背景

做异常值检测，数据有多个特征，执行完之后对每个数据基于密度计算一个score，然后对score进行排序，排序靠前的就认为是正常点，靠后的认为是异常点。

实验指标

指标主要是roc-auc和p@n。
roc-auc取值0-1,判断的是不同的正常/异常阈值下的排序能力，即score对与正常和异常数据的区分程度。
p@n的取值也是0-1,判断的是若取前N个为正常数据（N是实际的正常数据个数)的判断率，这个值越大说明结果判断的越正确

AUC 值	模型性能
1.0	完美分类器（所有异常得分 > 所有正常得分）
> 0.9	优秀
0.8 ~ 0.9	良好
0.7 ~ 0.8	一般
0.5 ~ 0.7	较差（接近随机）
0.5	随机猜测（模型无用）
< 0.5	比随机还差（可能是异常得分越低越异常，需取反）

数据集

数据集分为生成数据集和真实数据集。
生成数据集采用高斯分布在给定质点附近生成正常数据，再随机生成异常数据，来看不同算法的一个表现。gen_three_blob_data.py就是生成三个真实聚类的数据生成器，data里的threeblob_X.npy和threeblob_label.npy就是生成的500个数据。
真实数据集：采用真实数据经过去除缺失值和特征标准化处理
手写数据：[pendigits](#)

```
=====
📁 Pendigits 数据集 (ARFF 二分类版本)
=====

◆ 特征数 (d):          16
◆ 总数据量 (n):         10992
◆ 正常样本大小 (P):      9848
◆ 异常样本大小 (N):      1144
◆ 异常比例:             0.104 (10.4%)
✅ 已保存: .\pendigits_X.npy
✅ 已保存: .\pendigits_y.npy
=====
```

演讲数据: [speech\](#)

```
=====
📁 speech 数据集 (ARFF 二分类版本)
=====

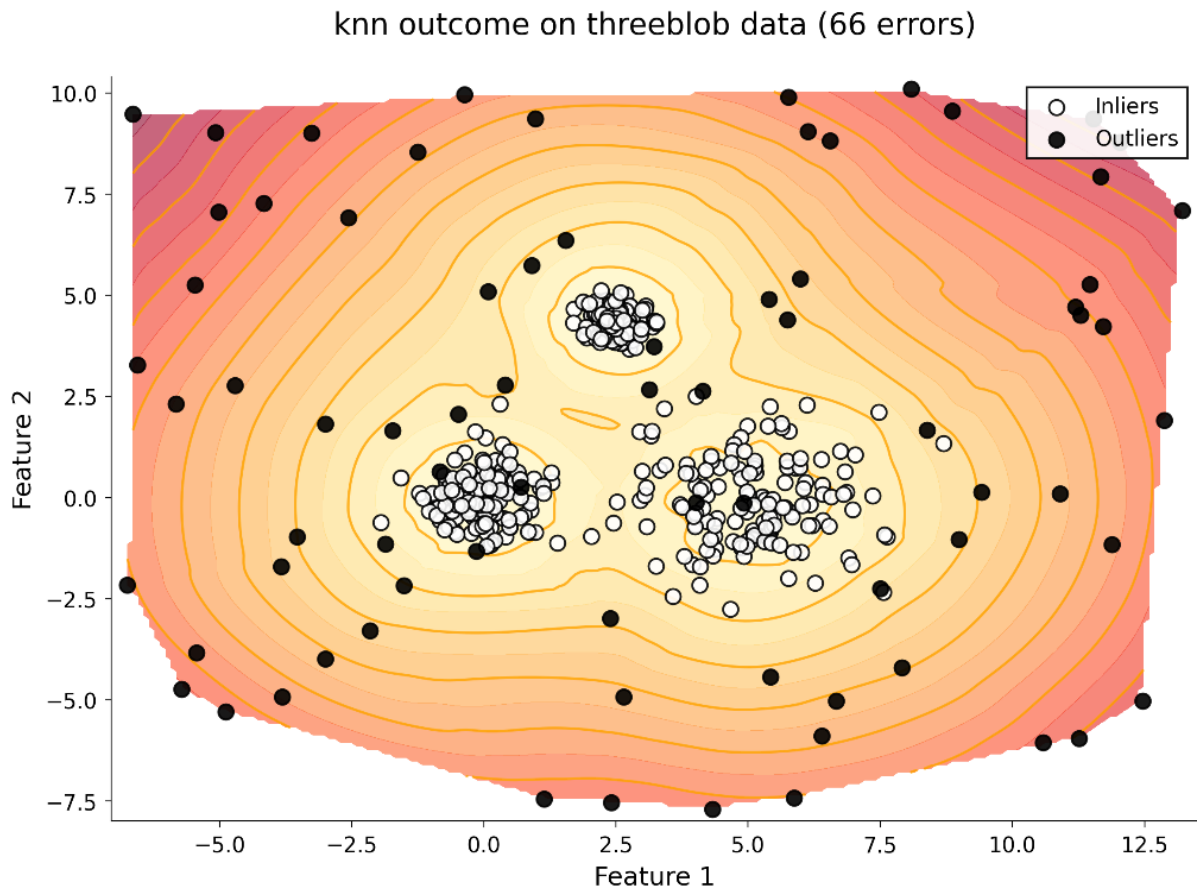
◆ 特征数 (d):          399
◆ 总数据量 (n):         3686
◆ 正常样本大小 (P):      3625
◆ 异常样本大小 (N):      61
◆ 异常比例:             0.017 (1.7%)
✅ 已保存: .\speech.npy
✅ 已保存: .\speech_y.npy
=====
```

模型

已实现baseline knn，还需要实现论文中的模型

输出

对于二维的生成数据，采用对score画热力图的方式展示结果。\\



\\

对于所有数据，生成roc和p@n指标的表格\\

```
=====
📊 ROC-AUC 表格
=====
Dataset  threeblob  pendigits  speech
Model
knn      0.972      0.324     0.479

=====
📊 Precision @ N 表格
=====
Dataset  threeblob  pendigits  speech
Model
knn      0.88      0.015     0.033
```