张涵　　PB20010469

lab4 实验报告

一. 实验目的

完成 DPC 算法的代码实现，并在给定数据集上进行可视化实验

二. 算法思想

1. Hyperparameter: a distance threshold d_c
2. For each data point i, compute two quantities:
   - Local density: $\rho_i = \sum_j \chi(d_{ij} - d_c)$,where $\chi(x) = 1$ if $x < 0$ and $\chi(x) = 0$ otherwise
   - Distance from points of higher density: $\delta_i = \min_{j:\rho_j > \rho_i} d_{ij}$ · For the point with highest density, take $\delta_i = \max_j d_{ij}$
3. Identify the cluster centers and out-of-distribution (OOD) points

- Cluster centers: with both high ρ_i and δ_i
- OOD points: with high δ_i but low ρ_i
- Draw a decision graph, and make decisions manually

三. 实验步骤

1. 读取数据集，（如有必要）对数据进行预处理
2. 实现 DPC 算法，计算数据点的 δ 和 ρ
3. 画出决策图决策图，选择样本中心和异常点
4. 确定分簇结果，计算评价指标评价指标，画出可视化图

四. 实验结果

具体可查看代码文件。