

实验 8_任务 2_二分类

问题：

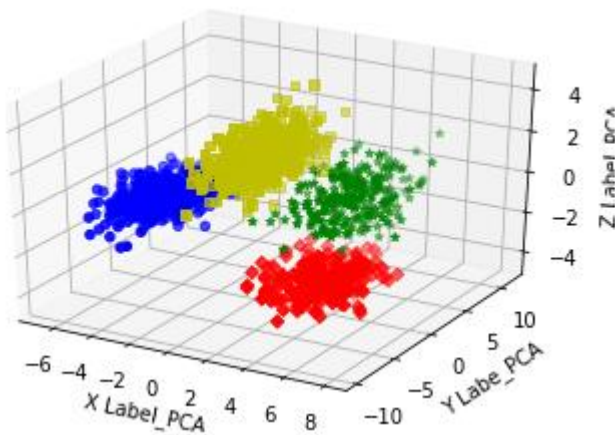
任务描述: 有 1500 个样例数据（每一行为一个样例，每个样例中有十一列数据，其中第一列为样例的类别，共四种 A、B、C、D，其后十列为样例的输入特征）。分类任务是通过输入特征来预测样例的类别。请完成下列工作:

1. 用降维算法将数据降为 2 维或者 3 维，并以不同的颜色表示各类别进行可视化。
2. 尝试比较不同的降维算法（如自己实现的 PCA 算法、sklearn 中的 PCA 方法，以及其他可能的降维方法等）的结果差异，如通过可视化结果进行比较等。
3. （附加题）尝试对此数据进行分类（分类方法、评估方法自行选择），并报告分类结果。

解答及实验过程记录：

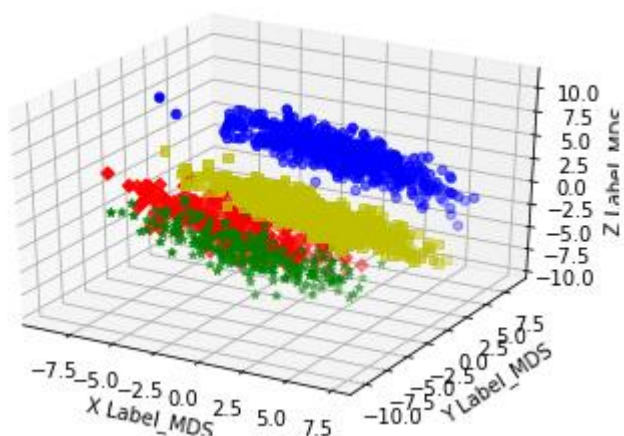
一 . 利用 PCA 降维

1.数据可视化：



二 . 利用 MDS 降维

1.数据可视化



问题 1：

首先，我用了 PCA 和 MDS 两种数据降维的方法进行了降维，并将降维后的结果进行了可视化分析，为了方便分析，我把数据集中的'A', 'B', 'C', 'D'分别用 1, 2, 3, 4 表示，并且在图中，分别用蓝，红，绿，黄表示 1, 2, 3, 4。

问题 2：

1. 观察上面的两种降维方法形成的数据的可视化结果，可以发现，四种类别的数据分别聚集在一起，由此可以得出结论：该数据集适合多分类训练。
2. 由于 PCA 和 MDS 两种降维方式本身的差异，可以发现在 3 维图中数据的分布特点仍然有差异。例如：PCA 的数据点主要沿 X 轴散开，沿 X 方向的数据离散程度最大；MDS 的数据点主要沿 Y 轴散开，沿 Y 方向数据的离散程度最大。
3. PCA 是主成分分析法，MDS 则较好的保持了原始数据的相对关系。