# 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。

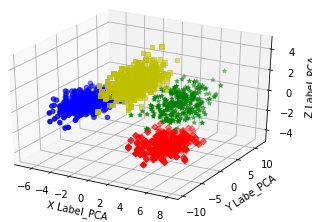
2. 尝试⽐较不同的降维算法（如⾃⼰实现的PCA算法、sklearn中的PCA⽅法，以及其他可能的降维⽅法等）的结果差异，如通过可视化结果进⾏⽐较等。

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：

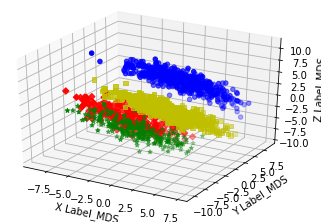
**一．利用PCA降维**

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。