**数据集竞赛**

1. **数据集说明**

本次竞赛包含两个数据集，两个数据集单独训练和测试，互不影响。

1. **Mnist手写数字数据集**

该数据集包含10个类别，分别是数字0~9。数据集被分为训练数据（traindata）和测试数据（testdata）两部分，其中训练数据包含1000张图像（每个类别100张图像），测试数据2000张图像（每个类别200张图像）。该数据集中的图像是二值图像，每个像素的取值为0或255，每张图像的分辨率为28\*28。

1. **Cifar彩色图像数据集**

该数据集包含10个类别，分别是飞机、汽车、鸟、猫、鹿、狗、青蛙、马、船和货车。数据集被分为训练数据（traindata）和测试数据（testdata）两部分，其中训练数据包含1000张图像（每个类别100张图像），测试数据1000张图像（每个类别100张图像）。该数据集中的图像是RGB三通道彩色图像，每张图像的分辨率为32\*32。

1. **竞赛任务及要求**
2. **竞赛任务**

对两个数据集单独处理。根据训练数据及相应的标签训练模型，用该模型对测试数据做分类，并统计测试数据的分类正确率。

1. **竞赛要求**

第一步，对所有图像提取特征。自由选择一种特征，如像素特征、颜色直方图特征等等。

第二步，用训练数据的特征和类别标签训练分类器。自由选择一种分类器，如K近邻、SVM分类器、朴素贝叶斯等等。

第三步，用训练好的模型对测试数据做分类，并统计测试数据的分类正确率。

注：两个数据集单独处理，可以采用不同的特征和分类器。在提取特征和训练分类器时，可以借助网络上的公开库函数。

1. **竞赛报告**

竞赛报告应包括两个数据集所采用的特征和分类器的介绍、带有注释的代码，以及实验结果分析。

1. **特别提示**

数据集竞赛的目的在于让同学们在实践中了解模式分类问题的求解方法和处理流程，只要符合要求，不管分类正确率高低均有成绩。**若发现代码抄袭，抄袭者与被抄袭者均判0分。**