**实验9 多分类服装识别实验**

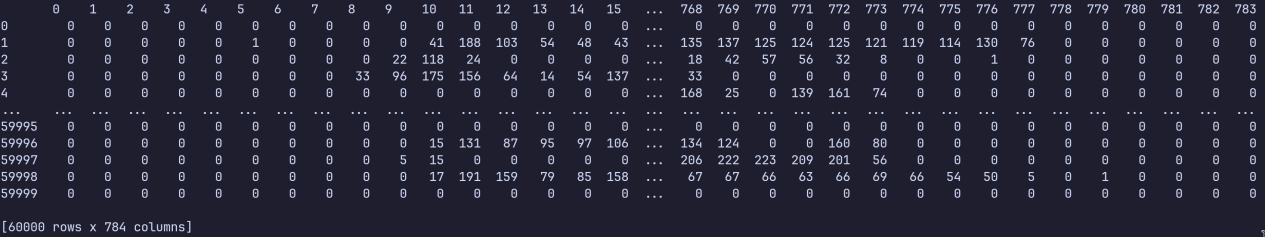
**1.实验数据**

**（1）训练集**

提供的数据包含60000张灰度图像，每张图像分辨率为28×28，存储于images\_train.csv文件中（无header行）。这些图像涵盖10种服装类型，分别用0 - 9表示。原始数据以CSV格式呈现，每行代表一幅图片，其数值为对应图片像素的灰度值。需注意，针对图像识别问题，通常需对不同图片的灰度值范围进行标准化处理，使每张图片的灰度值范围一致，而当前提供的数据尚未完成此步骤。训练数据集的标签存于labels\_train.csv文件中，用于标识每一行数据对应的服装类别。



images\_train.csv文件内容



1. **测试集**

另有10000条未知标签的测试数据图片，同样分辨率为28×28，保存在“images\_test.csv”文件里。每行数据为给定图片的灰度值，需通过建模识别每张图片所代表的服装类别。

1. **实验目的**
2. 设计有效的样本特征向量，锻炼特征工程能力。例如，可考虑运用PCA等方法对图片像素点的灰度值进行降维处理。
3. 借助机器学习分类算法，基于训练集训练出用于服装识别的分类器模型。
4. 将构建好的分类器模型应用于测试集，得出全体未知标签样本的分类结果。

**3.实验思路**

1. 附录中提供了可视化例程render.py，将该程序与数据文件images\_train.csv、标签文件labels\_train.csv置于同一文件夹下进行测试，可从中学习数据读取方法。另外，render\_teset.py用于测试数据的可视化。
2. 不限定具体的机器学习算法，以达到最佳预测效果为目标，追求尽可能高的准确率。可尝试集成多种学习模型。
3. 对所给图像的灰度值进行数据变换、降维等特征工程相关预处理，实现方法不做限制。
4. 自行学习多分类的分类器，如KNN、GNB、Logistic Regression、决策树、svm的SVC（from sklearn.svm import SVC）等。

**4. 实验要求**

1. 将预测结果保存至名为“preds.txt”的文本文件，文件内容为10000行，每行仅包含0 - 9中的一个数字，代表算法对测试数据的预测结果。预测数据顺序必须与测试集“images\_test.csv”中的样本顺序一致。
2. 上传两个文件：结果文件“preds.txt”和代码打包src.zip，分别以附件形式提交至学习通，无需提交本次实验报告文件。
3. 本次实验成绩评定采用竞赛机制。鉴于本次实验为多分类问题，将计算每位同学预测结果的准确率，并按从高到低的顺序评定相应的实验成绩。其中，准确率(accuracy)指分类正确的测试样本数占总测试样本数的比例。