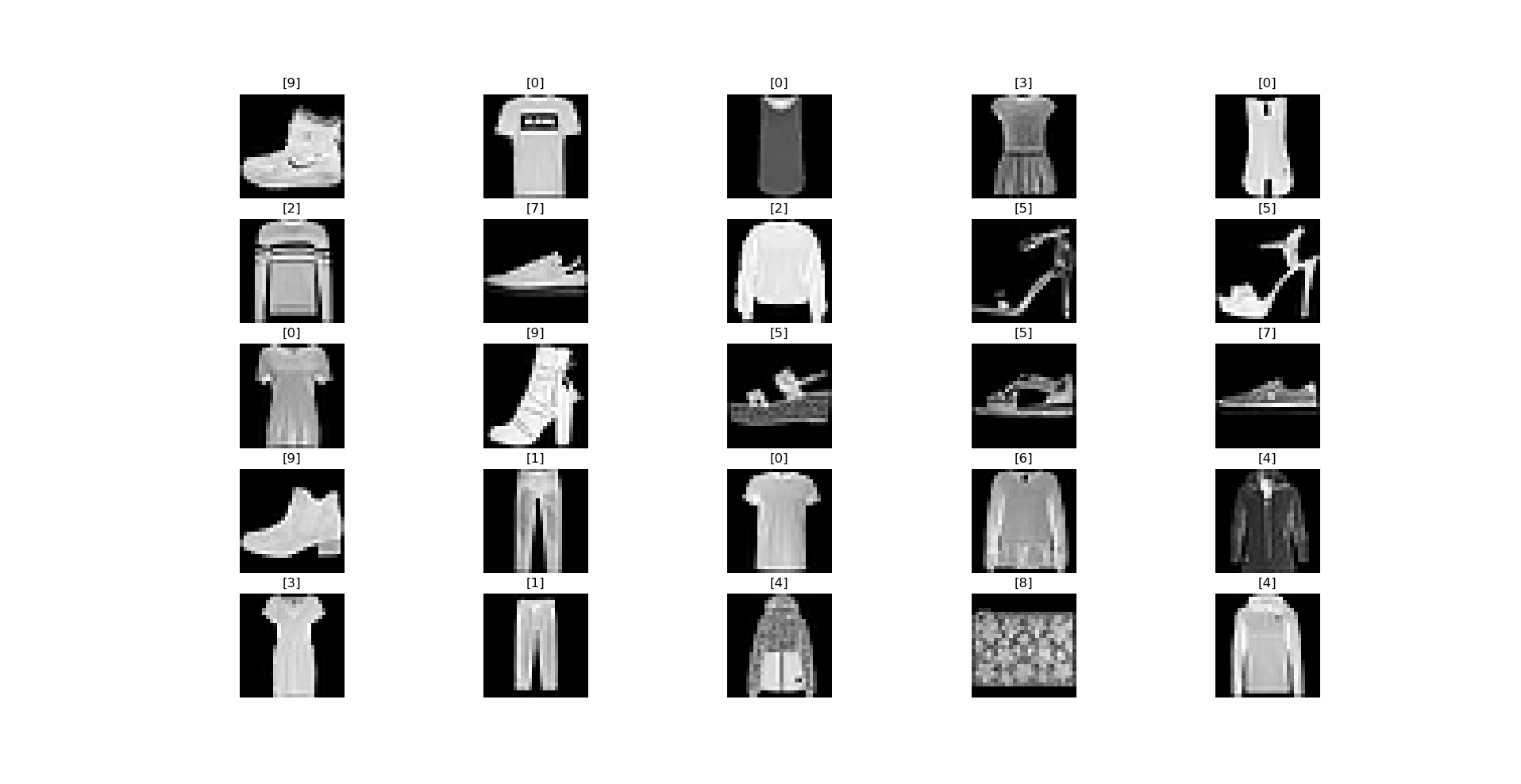
**实验9 多分类服装识别实验**

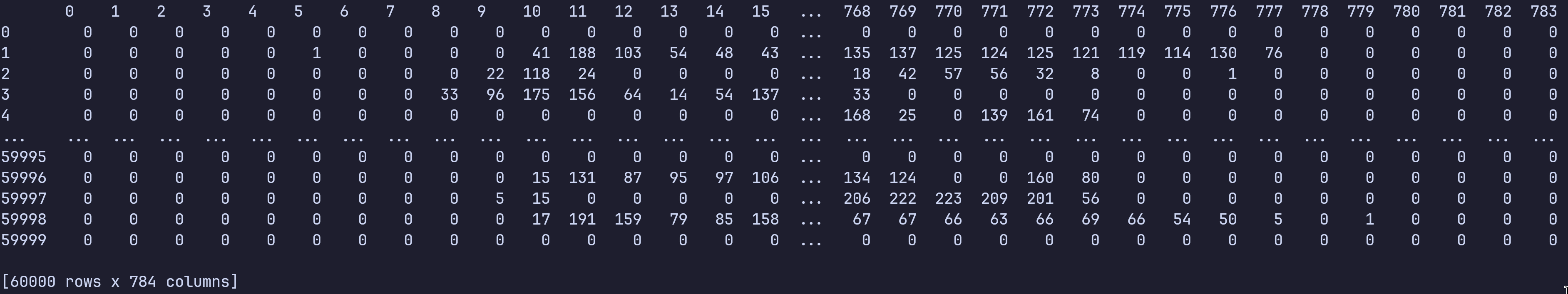
**1.实验数据**

**（1）训练集**

所给数据一共有60000张灰度图像（分辨率为28\*28），以images\_train.csv文件给出(没有header). 图像内容涵盖了10种服装，分别以0-9表示。 图像示例如图所示：



images\_train.csv文件内容如图



（训练集文件数据，其中的值是像素灰度值）

原始数据以CSV格式给出，每一行为一幅图片，为该幅图片的像素灰度值。请注意，一般就识别问题而言，对于图像可能需要将不同图片的灰度值范围做标准化处理，如使得每一张图片的灰度值范围都相同。而我们所给数据尚未完成这个步骤。

训练数据集的标签在labels\_train.csv中，表示每一行数据的标签。

**（2）测试集**

另有10000条未知标签的测试数据图片（分辨率也为28\*28），保存在“images\_test.csv”文件中。每一行为给定图片的灰度值，而每张图片代表的数值待建模识别。

**2.实验目的**

(1)设计样本的特征向量，具备一定的特征工程能力。如可以考虑对图片像素点的灰度值进行降维等处理（PCA等）；

(2)利用机器学习分类算法，基于训练集训练出服装识别的分类器模型；

(3)进而将构建好的分类器模型应用于测试集，给出全体未知标签样本的分类结果。

**3.实验思路**

(1)在附录中给出了可视化的一个例程（render.py，请把该程序和数据文件images\_train

.csv、标签文件labels\_train.csv放在同一文件夹下测试），可以从该程序中学习下数据的读取方法。另有render\_teset.py是对测试数据的可视化。

(2)具体的机器学习算法不限，以预测效果最佳为目标，追求准确率越高越好；可以尝试多种学习模型的集成。

(3)针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，实现方法不限。

(4)多分类的分类器请自学，如可以采用KNN，GNB，Logistic Regression，决策树，svm的SVC（from sklearn.svm import SVC）等

**4. 实验要求**

(1)将预测结果保存在名为“preds.txt”的文本文件中，内容为10000行， 每一行只有0-9数字中的一个，代表你的算法对测试数据的预测结果。预测数据顺序须与测试集“images\_test.csv”中的样本顺序保持一致。

(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。由于本次实验为多分类问题，我们将计算每位同学预测结果的准确率，然后由高到低进行排序评定相应的实验成绩。

**注：**准确率(accuracy)是指分类正确的测试样本数占总测试样本数的比例。