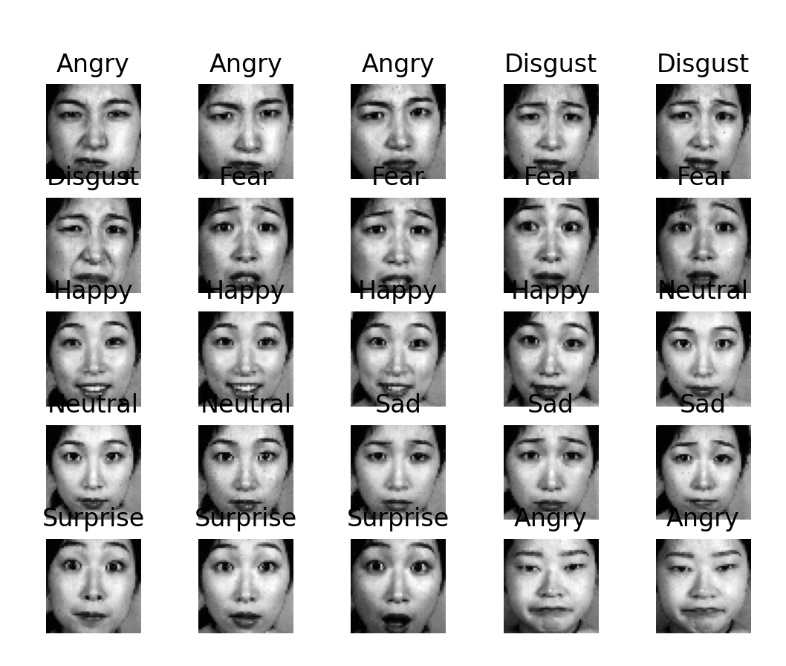
**实验9 多分类预测（表情识别）**

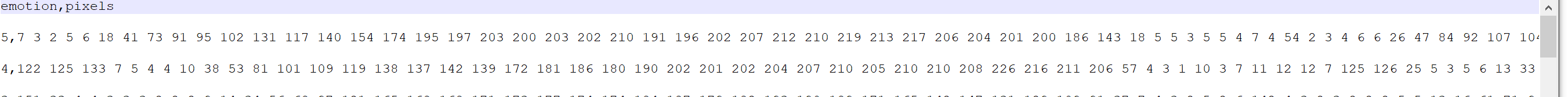
**1.实验数据**

**（1）训练集**

所给数据一共有170张灰度图像（分辨率为48\*48），目前以train\_data.csv文件给出. 图像内容涵盖了10个人， 每个人做出7种表情．7种表情包括： Angry，Disgust，Fear，Happy，Sad，Surprise，Neutral．（愤怒，厌恶，恐惧，高兴，悲伤，惊讶，中性），代码对应0-6。 图像示例如图所示：



train\_data.csv文件前两行如图



（训练集文件前两行数据，如第一行表情代码为5，后面是像素灰度值）

原始数据以CSV给出，每一行为一幅图片，其中第一列为表情代码，其余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。

**（2）测试集**

另有43条未知标签的测试数据图片（分辨率也为48\*48），保存在“test\_data.csv”文件中。每一行为给定图片的灰度值，而每张图片的表情类别(同训练集的代码0-6，含义相同)未知，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，基于训练集训练出表情识别的分类器模型；

(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train\_data

.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和归一化方法。

(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、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，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43行， 每一行只有0-6数字中的一个数字，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“test\_data.csv”中的样本顺序保持一致。

(2) 将结果文件“preds.txt”以附件形式提交至学习通。另请注意：只需提交结果文件“preds.txt”，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，即由于为多分类问题。我们将计算每位同学预测结果的准确率，，以然后由高到低进行排序评定相应的实验成绩。

注：准确率是指分类正确的测试样本数占总测试样本数的比例。