数据集

将处理好的蒙古字母图像整理到数据集中，数据集一共有三种，分别是训练集，验证集和测试集。数据集是基于23个手写蒙古字母图像通过裁剪、旋转组成的。数据集中每个图像的尺寸是32\*32\*1。

其中训练集每个手写蒙古字母有100张图像，一共有2300张图像。

其中验证集每个手写蒙古字母有10张图像，一共有230张图像。

其中测试集每个手写蒙古字母有1张图像，一共有23张图像。

读入图像

读入训练集以及验证集图像的操作分为两步：

第一步：遍历图像目录获取图像个数，具体操作过程为：通过os.walk(dir)与for循环搭配使用，遍历数据集的每一个图像，获取图像总个数记为count。设置存储图像信息images、图像标签labels两个二维数组，根据count的数量、图像的尺寸SIZE的大小以及标签CLASSES的个数，初始化这两个二维数组，代码如下：

images = np.array([[0] \* SIZE for i in range(count)])

labels = np.array([[0] \* CLASSES for i in range(count)])

第二步：遍历图像目录生成图像数据和标签，具体操作过程为：通过调用os.walk(dir)，遍历数据集的每一个图像，用index记录图像的标签，并循环遍历图像。通过img.getpixel((w, h))获取每个图像的像素点，让每个图像的像素点与阀值进行比较，将结果存入images[index][w + h \* width]二维数组中，代码如下：

if img.getpixel((w, h)) > 128:

images[index][w + h \* width] = 1

else:

images[index][w + h \* width] = 0

其中：images二维数组中每行代表一张图像的信息，其中1表示有效信息。将标签存入labels[index][i]二维数组中，每行代表图像的标签，其中1表示这个图像对应的标签是这个列的数值。