**NMDA 2021 第三次作业**

MNIST是视觉处理方面的结构比较简单的标准数据集，其中包括黑白手写数字0-9图像，每张图像有对应的标注。原始数据集的连接<http://yann.lecun.com/exdb/mnist/>。

我们已经将数据集整理，包括手写16\*16像素的黑白手写数字图像和对应的数字标记，并分为train，validation和test三组。图像数据已经转换为256长度的向量。

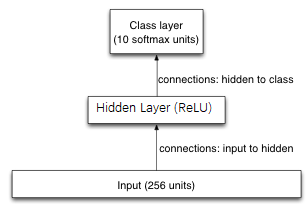
本次作业中，要求使用numpy，完成Network.py中的Network类和Assignment3.ipynb，最终实现包含一层隐层的神经网络，并在MNIST数据集上进行分类任务。请根据Assignment3.ipynb的引导，完成对应的任务。numpy的使用可以参考官方文档<https://numpy.org/doc/stable/>。

Assignment3.ipynb将引导你完成任务。神经网络类Network位于Network.py中。Network类中定义了多个反向传播算法相关的方法，目前只是定义了接口，在下面的任务中需要完成代码。

**本次作业最终要求提交Network.py，Assignment3.ipynb和反向传播推倒过程back\_prop.pdf。请不要提交数据文件。**

**任务1（20 mark）**

我们要实现一个包含隐层的神经网络。如图所示：



输入层x具有256个单元，对应图片中的像素点。隐层h单元数量由超参数决定。分类层c具有10个单元，对应0-9十种类别。激活函数选择ReLU。经过分类层通过softmax层后使用交叉熵得到网络的损失函数。网络的损失函数加L2正则化后得到最终的损失函数。即：

其中，x为输入的图像数据，y为label。

**任务1中，要求完成参数的初始化和前向传播过程**，对应Network类的\_\_init\_\_方法和forward\_pass方法。

如果你对python中面向对象概念不熟悉，可以参考<https://zhuanlan.zhihu.com/p/37145434>。

1. 参数初始化要求W1，W2初始化为标准差std的正态分布，b1，b2初始化为0。
2. 前向传播方法具有缺省参数y=None，要求输入标签y时返回loss，不输入标签y时返回预测的标签predict。

**任务2（20 mark）**

推导反向传播过程的梯度，得到Loss分别对W1，b1，W2，b2的导数。推导过程请以back\_prop.pdf提交。推荐使用LaTeX，也可以使用手写+拍照。

**任务3（20 mark）**

为检验反向传播算法的正确性，要求使用数值计算梯度进行对比。反向传播和数值计算梯度分别对应back\_prop方法和numerical\_gradient方法。需要返回grads，类似于params成员，其中包括W1,b1,W2,b2所对应的梯度。

数值计算梯度时可以使用差分法，先定义出一个较小的delta，使用

对梯度进行数值计算。

1. **完成反向传播计算**
2. **完成数值梯度计算，检验正确性**

**任务4（15 mark）**

得到了正确的梯度后，可以开始使用随机梯度下降方法进行网络训练，这部分对应train方法。请使用之前实现的前向传播算法和反向传播算法，实现支持momentum的随机梯度下降。在随机梯度下降迭代过程中，每次需要选取数据集的一部分(minibatch)作为数据。为了查看训练过程，需要记录下训练中的loss和正确率的变化过程。代码中提供了4个容器用来记录loss\_history, acc\_history, val\_loss\_history, val\_acc\_history。可以使用append添加记录。

1. 实现梯度下降，支持momentum和minibatch
2. 记录训练过程中的loss变化和正确率变化

**任务5（10 mark）增加训练策略**

我们已经得到了一个可以工作的神经网络，并且实现了使用梯度下降进行训练。下面，我们使用一个256隐层的网络去训练，观察训练结果。

此后，增加两种训练策略：Early stop策略是一种简单的防止过拟合的策略，通过验证集上loss的变化监测何时停止训练。 Learning rate decay 是另一个常见的训练策略，通过逐步减小的learning rate达到更好的训练效果。一种常见的做法是，每经过一定次数的迭代后将learning rate乘一个衰减系数。

1）实现early stop

2）实现learning rate decay

**任务6（15 mark）**

调整超参数，得到尽量好的神经网络。初始化参数和网络参数都可能对最终的训练结果产生影响。

网格搜索法是一种常见的调参方法。将不同参数的可能取值列出，选择参数取值不同排列组合进行训练，最后可以选出最优的一组参数。

随机搜索法是另一种常见的调参方法。当网格搜索法调参需要的计算量过大时，可以使用随机搜索法进行，通过不断的随机出超参数进行训练，多次随机后选择最优的一组参数。

1）使用网格搜索法或随机搜索法，得到一组超参数。除了列出的超参数外，也可以选择调整别的参数。请保留你的调参过程的代码。

2）展示得到的超参数进行训练的结果，在notebook中记录选择的参数和在测试集上得到的正确率。