

神经网络课程作业 4

1. Batch normalization 的输出服从什么样的分布（指出分布中的具体参数）。
2. Batch normalization 为什么归一化后还有放缩（ γ ）和平移（ β ）？
3. 目前有一批病人的身体数据（体重变化，血液指标等）和他们是否患有肺癌的真实标签，其中患肺癌的样本只占非常小的比例。数据直接送入一个神经网络中，求问应该使用什么样的初始化？数据中不同的特征数值差异过大，求问如何改进能够让网络更好地学习数据中的分布？
4. 参考以下代码使用 `scikit-learn` 包获取波士顿住房数据集（如果无法正确获取，请检查 `scikit-learn` 的版本）。

```
# 导入数据集
from sklearn.datasets import load_boston
boston = load_boston()

# 定义网络模型
class MLP(nn.Module):
    def __init__(self):
        super(MLP, self).__init__()
        # First hidden layer
        self.h1 = nn.Linear(in_features=13, out_features=20, bias=True)
        self.a1 = nn.ReLU()
        # Second hidden layer
        self.h2 = nn.Linear(in_features=20, out_features=10)
        self.a2 = nn.ReLU()
        # regression predict layer
        self.regression = nn.Linear(in_features=10, out_features=1)

    def forward(self, x):
        x = self.h1(x)
        x = self.a1(x)
        x = self.h2(x)
        x = self.a2(x)
        output = self.regression(x)
        return output
```

结合上述代码完成：

- a) 请把获取的数据集分为训练集和验证集，它们的比例为 7:3，设计一个三层的神经网络（不建议调用 `MLPRegressor` 函数，可以使用 `PyTorch` 框架自行搭建）分别在进行了归一化和不对数据进行归一化的情况下进行训练，给出训练集 `loss` 的折线图，以及验证集真实值和预测值的差异图，比较二者之间的区别。（推荐 `batch size` 设置为 8，迭代次数为 5，优化器使用 `SGD` 优化器，学习率为 0.001，损失函数为平方误差损失）
- b) 分别使用平方误差损失，绝对值误差损失和 `Huber` 损失作为损失函数进行训练，给出训练集 `loss` 的折线图，以及验证集真实值和预测值的差异图，比较彼此之间的区别。
- c) 改变不同的学习率（其余参数固定），给出训练集 `loss` 的折线图，以及

验证集真实值和预测值的差异图。再比较固定学习率和学习率衰减二者之间的区别，给出训练集 loss 的折线图，以及验证集真实值和预测值的差异图。

5. 参考下列代码下载 mnist 数据集并使用一个两层的神经网络进行分类（实例代码如下），请分别画出训练集准确率，训练集 loss 及验证集准确率的折线图。（推荐 batchsize 设置为 100，迭代次数为 3，优化器使用 SGD 优化器，学习率为 0.001，损失函数为交叉熵损失）

```
# 载入数据集
train_data=torchvision.datasets.MNIST(
    root='MNIST',
    train=True,
    transform=torchvision.transforms.ToTensor(),
    download=True
)
test_data=torchvision.datasets.MNIST(
    root='MNIST',
    train=False,
    transform=torchvision.transforms.ToTensor(),
    download=True
)

# 网络模型
class MLP(nn.Module):
    def __init__(self,):
        super(MLP, self).__init__()
        # 第一个线性层
        self.layer1=nn.Linear(784,2048)
        self.layer2=nn.Linear(2048,10)
        self.relu=nn.ReLU()

    # 前向传播
    def forward(self,input):
        out=self.layer1(input)
        out=self.relu(out)
        out=self.layer2(out)
        return out
```

结合上述代码完成：

- 针对上述神经网络，请至少使用三种不同的初始化方式，并分别画出训练集准确率，训练集 loss 及验证集准确率的折线图。
- 针对上述神经网络，请分别进行批归一化，层归一化，实例归一化及组归一化，并画出训练集准确率，训练集 loss 及验证集准确率的折线图。

注意：对于 4、5 题，请在 pdf 文件中贴关键代码和实验结果，并配以必要的注释。