

第二次作业



■ 理论题：25%

1. 单热向量是指假设共有 C 类，样本 x_i 属于第 j 类，则样本 x_i 的标签对应的单热向量编码为单位向量 $\mathbf{y}=(0, \dots, 0, 1, 0, \dots)$ ，其中 \mathbf{y} 在第 j 个位置为1，其他全0。现假设样本来自三个类，某次训练中的一个batch包含3个训练样本 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ ，分别来自第1, 2, 3类

a) 试推导采用单热向量编码时该batch交叉熵损失函数表达式。（提示：设该batch对应网络输出为 $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ ）

b) 如果网络输出为 $\mathbf{y}_1=(0.65, 0.43, 0.11)$, $\mathbf{y}_2=(0.05, 0.51, 0.18)$, $\mathbf{y}_3=(0.33, 0.21, 0.72)$ ，计算交叉熵损失函数值。

2. 令 $\mathbf{y}=\text{sigmoid}(\mathbf{W}^T\mathbf{x}+\mathbf{b})$ ，求采用均方误差函数 $J=\|\mathbf{y}-\mathbf{z}\|^2$ 时 J 对 \mathbf{W} 和 \mathbf{b} 的具体导数表达式，写出详细的推导过程和完整的表达式

矩阵求导可参看：<http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html>

第二次作业



- 实践题：75%=35%+30%

- 1) 实现一个三层神经网络，并使用iris数据集前80%训练、后20%测试，要求测试误差小于5%，分析至少三种非线性激活函数的影响。
- 2) 设计并实现一个深度学习网络结构，能够在MNIST数据集上 (前6万个训练，后1万个测试)获得至少99%的测试精度

- 附加题：(可不做，做对20%额外分)

仅使用numpy实现三层神经网络BP训练算法(输入d维向量，中间h个隐含神经元，输出c>1类（单热向量编码），隐含层使用sigmoid激活函数，输入输出层使用线性激活函数，损失函数为均方误差或者交叉熵，并在iris数据集上测试，结果与实践题1)进行比较

- 注意：实践题1)和2)可以使用tensorflow或pytorch或其他深度学习软件包，要求绘制出训练过程中训练数据和测试数据的损失函数值，训练精度和测试精度变化曲线
- 发现作弊或抄袭，本次作业0分处理

作业提交



- 文档：包括理论题的回答和实践题的网络结构说明，结果截图
- 代码：包括所有运行需要的代码和数据
- 把以上两部分压缩在一个压缩包，发送到邮箱wangzihao33@sjtu.edu.cn
- 邮件标题格式：课程ML作业2_学号_姓名
- DDL：第9周周日（11月11日）23:59之前（以邮件收到时间为准）

MNIST和Iris读取代码



- `import scipy.io as sio`
- `data_dict = sio.loadmat('iris.mat')`
- `samples = data_dict['samples']`
- `labels = data_dict['labels']`

- `data_dict = sio.loadmat('MNIST.mat')`
- `samples = data_dict['fea']`
- `labels = data_dict['gnd']`
- `x=samples[0].reshape(28,28) # 变换成28x28的矩阵`
- `plt.imshow(x) # 显示图像`