

作业



- 理论题：50%=25%+25%

1. 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. 假设输入有2个样本 x_1, x_2 : (1)请画计算 x_1, x_2 标准差的详细计算图; (2) 标出当 $x_1=-1, x_2=3$ 时输出对图中每个节点输入变量的梯度值, 并求出 x_1, x_2 总的梯度值

作业



■ 实践题一：25%

- 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，要求测试错误率小于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

■ 实践题二：25%

1) 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)，损失函数用均方误差或者交叉熵

2) 在iris数据集上对1)中实现的算法测试，并与实践题一的结果进行比较

■ **注意：**实践题一可以使用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曲线，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：`ymsun_sjtu@sjtu.edu.cn` (IEEE班), `tangshuai@sjtu.edu.cn` (AI班)
- 邮件标题格式：课程MLKD作业2_学号_姓名
- DDL：第9周周日（11月8日）23:59之前（以邮件收到时间为准）