

提醒注意：

- 本次作业截止于2025年5月9日。
- 作业分为三部分：问答题、实训题、以及实训题报告
 - 问答题答案可以手写并扫描，或者用latex（或word）手打，最终以QA.pdf文件命名。
 - 实训题按照项目共享链接内要求和基础代码进行作答。
 - 报告部分同样可以手写或者手打，以Report.pdf文件命名。
 - 作业提交格式：< studentID >_< name >_A2.zip。比如1921102_田嘉怡_A2.zip
 - 提交的zip文件要求（仅）包括：
 - * 实训题文件：包括 main.ipynb，以及predict.csv文件。
 - * 问答题答案：QA.pdf
 - * 报告：Report.pdf
- 作业压缩包需要在spoc平台上提交。
- 每迟交1天（不满1天按1天计算），本次作业扣除10%分数。
- 不按作业要求和格式提交，视情况扣分。不得抄袭。

第一部分：问答题（4分）

Q 1 神经网络基础 (1分)

围绕神经网络基础，回答以下问题：

1. 给定一个3层全连接网络（输入层2节点，隐藏层3节点，输出层1节点），写出前向传播的矩阵运算形式（使用Sigmoid激活）。
2. 交叉熵损失函数如何衡量预测概率分布与真实分布的差异？写出二分类问题的交叉熵公式。

Q 2 模型优化 (1分)

围绕深度学习模型优化，回答以下问题：

1. 在前馈神经网络中，所有的参数能否被初始化为0？如果不能，能否全部初始化为其他相同的值？原因是什么？
2. 给定损失函数 $L(w) = w^2 + 2w + 1$ ，计算权重 w 在初始值 $w_0 = 3$ 时的梯度，并说明梯度下降的更新方向。

Q 3 激活函数 (2分)

logistic sigmoid函数在深度学习中用途很广。根据函数表达式回答下列问题：

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

1. 证明 $1 - \sigma(x) = \sigma(-x)$ ；
2. 证明 $\sigma'(x) = \sigma(x)(1 - \sigma(x))$ ，其中 $\sigma'(x) = \frac{d}{dx}\sigma(x)$ 。并画出 $\sigma(x)$ 和 $\sigma'(x)$ 的函数图像；

3. 神经网络通常都是layer-by-layer处理，比如，对于输入 \mathbf{x} ,输出 \mathbf{y} 可以用下列式子表示：

$$\mathbf{y} = f^{(L)}(f^{(L-1)}(f^2(f^1(\mathbf{x})))) \quad (2)$$

式子中 $f^{(i)}$ 代表着神经网络中的第 i 层，当神经网络层数（式子中的 L ）很大的时候，我们称为深度神经网络。神经网络优化算法一般选择随机梯度下降算法，我们用 $\boldsymbol{\theta}$ 表示神经网络中的参数， \mathbf{g} 表示神经网络对应损失函数的梯度，神经网络参数更新用下式表示：

$$\boldsymbol{\theta}^{new} \leftarrow \boldsymbol{\theta} - \lambda \mathbf{g} \quad (3)$$

式子中 λ 是学习率。梯度计算采用链式法则。设 $\boldsymbol{\theta}^{(i)}$ 为第 i 层的参数， $\mathbf{y}^{(i)}$ 是 i 层的输出：

$$\frac{\partial l}{\partial (\boldsymbol{\theta}^{(i)})^T} = \frac{\partial l}{\partial (\mathbf{y}^{(i)})^T} \frac{\partial \mathbf{y}^{(i)}}{\partial (\boldsymbol{\theta}^{(i)})^T} \quad (4)$$

式子中 l 是损失函数，我们的目标是将损失函数最小化。计算过程被称为损失函数的反向传播。因为在这个过程中，损失函数从最后一层传播到第一层。这种方法经常会遇到梯度消失问题，这意味着，梯度 $\frac{\partial l}{\partial (\boldsymbol{\theta}^{(i)})^T}$ 变得非常小，即当梯度从最后一层传播到第一层时， $\|\frac{\partial l}{\partial (\boldsymbol{\theta}^{(i)})^T}\| \rightarrow 0$ （很快趋向0）。sigmoid激活函数在神经网络中是一种常用的激活函数，即每一层的函数 $f^{(i)}$ 均对其输入进行逐元素的sigmoid激活。

证明sigmoid激活函数很容易导致梯度消失问题。（提示：你可以观察在梯度下降过程中网络的某一个元素，也可以参考（2）画出的函数图像。）

第二部分：实训题（共4分）

实训题要求：

- 本次作业包括1个实训题，作业要求以及基础代码以Aistudio项目的形式发布。
- 发布项目链接有效期3天，请在作业发布3天内fork这个项目，生成“我的项目”，并在自己fork的项目下进行作答，生成答案后按要求保存提交。

Q 1 动物图片分类-构造神经网络

在本作业中，首先给出了手写数字识别的示例。该示例中，展示了数据准备，网络配置和模型训练评估三部分的内容。阅读该示例可以帮助理解深度学习的一般过程。在动物图片分类的项目中，有以下任务需要完成，助教将根据任务完成的情况来给予分数。评判标准包括代码和实验结果两部分，实训题给分更关注代码逻辑和完整性。请在实验报告中写明实验结果和代码思路。

助教已经构建好了一个简单的卷积神经网络和训练流程，这是代码中已经实现的部分，你可以在这基础上进行改进。以下是对于改进的一些提示：

- 采用数据增强（data augmentation），对输入图片进行裁剪或者旋转等来提高性能。
- 改变网络结构，例如增加卷积层数，或者使用VGG等其他深度学习网络。注意：不要采用任何的预训练模型。（复杂的网络有过拟合风险，调整网络结构后超参需要随之调整。）
- 借助无标注数据，引入半监督学习（伪标签），观察无标注数据对分类性能的影响。
- 理解和补全ViT（Vision Transformer）模型，并分析实验结果。

实验介绍详情和参考基础代码请参见Aistudio中的共享项目“人工智能作业二-动物图像分类”。

第三部分：实训题实验报告（共2分）

- 请按照实验报告模板完成实验报告。
- 实验报告模板是通用模板，可根据每个作业要求的差别，自由进行微调。