

提醒注意：

- 本次作业截止于2025年4月10日。
- 作业一分为三部分：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 >_A1.zip。比如ZY1921102_田嘉怡_A1.zip
 - 提交的zip文件要求（仅）包括：
 - * 实训题文件：包括 todo.py 和 predictions.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（共4分，每题1分）

Q 1（线性回归与线性分类）

请回答以下关于线性模型 $f(x) = \mathbf{w}^T \mathbf{x}$ 的相关问题（偏置项 b 已融入进参数 \mathbf{w} 中）。

- 1) 在线性回归中，如果我们给每个样本 (\mathbf{x}_i, y_i) 赋予一个权重 r_i ，经验风险函数为 $L(\mathbf{w}) = \sum_{i=1}^n r_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2$ ，计算其最优参数 \mathbf{w}^* ，并分析权重 r_i 的作用。
- 2) 证明在二类线性分类中，权重向量 \mathbf{w} 与决策平面正交。
- 3) 在Logistic回归中，是否可以通过最小化 $(y - \text{sigmoid}(\mathbf{w}^T \mathbf{x}))^2$ 来优化参数 \mathbf{w} ？为什么？

Q 2（正则化）

加入L2正则化(regularization)后，线性回归的损失函数的标准形式为：

$$L = (Y - X\mathbf{w})^T (Y - X\mathbf{w}) + \lambda \mathbf{w}^T \mathbf{w}, \quad \lambda > 0$$

- 1) 假设L2正则化项被误写为 $\lambda Y^T Y$ ，请解释为什么该项起不到正则化的作用。
- 2) 假设L2正则化正确，但 λ 小于0，请解释为什么该项起不到正则化的作用。

Q 3（支持向量机）

考虑一个二维特征空间中的二分类问题。训练集包含8个样本，其在二维空间中的分布如图1所示。图1中不同类别的样本用不同的形状表示，根据图1回答以下问题。

- 1) 使用实线分类画出使用SVM分类器可得到的分类边界，并用圆圈标出支持向量
- 2) 在该分类边界下所得训练错误率为多少？

3) 移除那些样本可以让分类边界发生变化

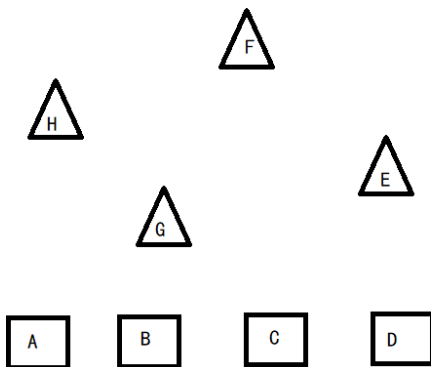


图 1: 两类分类问题中8个样本的分布图

Q 4 (决策树)

考虑下面如表1所示的一个数据集，它记录了某学生多次考试的情况，请根据提供的数据按要求构建决策树。

- 1) 根据信息增益率选择第一个属性，构建一个深度为1的决策树（根节点的深度为 1）。
- 2) 根据信息增益率构建完整的决策树。请回答，这两个决策树的决策结果是否和训练数据一致，并解释说明。

表 1: 某学生多次考试情况

是否通过考试	是否认真复习	是否超常发挥
是	是	否
是	是	是
是	是	否
是	是	是
是	是	否
是	否	是
否	否	否
否	否	是

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 二分类-垃圾短信分类识别

垃圾短信 (Spam Messages, SM) 是指未经过用户同意向用户发送不愿接收的商业广告或者不符合法律规范的短信。随着手机的普及，垃圾短信在日常生活日益泛滥，已经严重的影响到了人们的正常生活娱乐，乃至社会的稳定。

本实验要求使用Pandas, Numpy, Sklearn 等库进行相关特征处理，使用 paddle 框架训练分类器，使用机器学习的方法对垃圾短信进行分类。

实验介绍参见视频：实验说明视频，密码：BOGC

实验参考基础代码请参见Aistudio中的共享项目“人工智能课程-作业一-垃圾短信分类”。

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。