

作业答案

1.说明有监督学习和无监督学习的各自的特点及区别

- 监督学习
 - 有标签
 - 分类：标签值离散
 - 回归：标签值连续
- 无监督学习
 - 无标签，按照样本的相似性进行聚合
- 半监督学习
 - 部分有标签，部分无标签
- 强化学习

2.说明下机器学习的建模流程

- 获取数据
- 数据基本处理
- 特征工程
- 模型训练
- 模型评估

3.谈一下你对特征工程的理解

用专业背景知识和技巧处理数据，让机器学习算法效果最好

- 特征提取
- 特征预处理
- 特征降维
- 特征选择
- 特征组合

4.说下模型拟合问题及产生的原因

- 欠拟合:训练集上表现很差、在测试集表现也很差
 - 原因：模型过于简单
- 过拟合:训练集上表现很好、在测试集表现很差
 - 原因：模型太过于复杂