# 绪论

1.1 将音乐软件的推荐问题视作一个机器学习问题，如果用户喜欢这首音乐会点击“喜欢”按钮或收藏。请根据Tom Mitchell对机器学习的定义写出该问题的E、T、P分别代表什么？

参考答案：经验E是指推荐给用户的音乐与用户的反馈等历史数据，任务T是指推荐音乐这，P是指用户对推荐的音乐喜欢与否。

1.2 简述分类任务与回归任务的区别，并给出应用举例。

参考答案：输入变量与输出变量均为连续变量的预测问题是回归问题，如判断是否为垃圾邮件；输出变量为有限个离散变量的预测问题成为分类问题，如预测房价；

1.3 假设某一数据集有50条数据、三个属性值，每个属性值分别有3、4、6种取值，请计算该样本空间的大小。

参考答案：由题意，共有（3+1）\*（4+1）\*（6+1）+1=140种。

# 模型评估与选择

2.1 简述经验误差与泛化误差的区别。

参考答案：经验误差指学习器在训练集上的误差，泛化误差指的是学习器在新样本上的误差。

2.2 简述如何通过ROC和AUC这两个指标判定学习器的优劣。

参考答案：若一个学习器的ROC 曲线被另一个学习器的的曲线完全‘包住’，则后者的性能优于前者；若两曲线交叉，则无法判断；当两个学习器的ROC 曲线交叉时，通过比较AUC 来判断两个学习器的性能优劣。AUC 值越大，则学习器性能越好。

2.3

# 线性回归

# 决策树

# 神经网络

# 支持向量机

# 贝叶斯分类器

# 集成学习

# 聚类