

# 2023-2024 学年秋季学期《模式识别与机器学习》课程

## 第 12 章作业题

### 一、选择题

1. (多选题) 关于特征选择方法, 以下说法正确的是 ( )
  - A. 数据降维和特征选择都可以应对维度爆炸问题
  - B. 前向搜索、后向搜索和双向搜索都是贪心策略
  - C. 过滤式特征选择的计算开销通常比包裹式特征选择更大
  - D. 信息熵、类内类间距离、随机变量间的相关性都可以作为特征子集评价指标
2. (多选题) 以下关于特征选择的说法, 正确的是 ( )
  - A. 一般而言, 双向搜索策略能得到最优特征子集
  - B. 特征选择不仅可以降低特征空间的维度, 还可以消除特征之间的相关性
  - C. 可以将分类器错误率作为设定特征评价准则的依据
  - D. 特征数量太多会影响模型参数估计的稳定性

### 二、计算题

1. (特征子集搜索与评价) 现有 7 个二维特征向量:

$$\{(-7, 3.4), (-2, 3.4), (-0.5, 2), (0.5, 2), (2.5, 3.4), (3, 3.4), (3.5, 3.4)\},$$

且对应的标签分别为 $\{0, 0, 1, 0, 1, 1, 1\}$ 。记每个特征向量的两维特征分别为  $x$  特征和  $y$  特征, 请使用**基于类内类间距离的判据**作为评价指标, 判断  $x$  特征和  $y$  特征的优劣。