

考试科目：《机器学习与数据挖掘》（A 卷）

警示：考试作弊者，不授予学士学位！

一、简答题（共14小题，每小题5分，共70分）

- 1、描述什么是机器学习的泛化能力.
- 2、简述K-Nearest Neighbors (K-近邻)算法的原理.
- 3、以信息增益作为属性的划分方法，简述决策树的算法原理及构建过程.
- 4、简述决策树的预剪枝策略.
- 5、简述Bagging集成方法的原理并举例.
- 6、详细描述随机森林的原理及算法过程.
- 7、随机森林为什么不容易过拟合？
- 8、写出支持向量机基本模型的优化目标函数.
- 9、简述核函数的作用原理.
- 10、简述什么是自注意力（self-attention）.
- 11、简述特征选择与数据降维的相同及不同之处.
- 12、简述主成分分析（PCA）的原理.
- 13、详细描述DBSCAN算法流程.
- 14、解释Expectation Maximization（EM）算法的M-step原理.

二、思考题（共 3 小题，每小题 10 分，共 30 分）

- 1、描述基于内容（content-based）、协同过滤（collaborative filtering）、因式分解（factorization）的推荐系统方法，并对比三者的差异.
- 2、随机森林、支持向量机、深度神经网络是三类性能强大的机器学习模型，分析它们成功背后的原因.
- 3、你认为机器学习的价值和意义是什么.