



# 机器学习

## -Lab3

戴蕾

dailei@ecust.edu.cn

2024年春季

## ◆ 实验目标

- 掌握KNN的基本原理及python实现
- 掌握决策树的基本原理及python实现
- 运用KNN实现具体任务
- 运用决策树实现具体任务

## 1.2 练习内容1

### ➤ KNN

- 见学习通 “资料—**Lab3-KNN**”

名称

📁 练习1

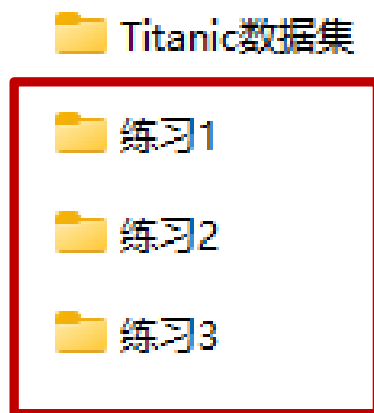
📁 练习2-海伦约会

📁 练习3-数字识别

## 1.2 练习内容2

### ➤ 决策树

- 见学习通 “资料—Lab3-决策树”



## 2.1 实验内容1

### ➤ KNN

#### ◆ 基本要求

任选**海伦约会或数字识别任务**，在其**code**基础上增加**交叉验证**等方法确定**最优K值**，最终需要输出**最优K值和对应精度**，并撰写**实验报告2**。

 练习2-海伦约会

 练习3-数字识别

## 2.2 作业提交

### ➤ 提交形式

#### □ 两个压缩包:

- 压缩包1: 实验报告本 (word+pdf)
- 压缩包2: 完整可编译代码 (包括数据集)

#### □ 命名方式: 学号\_姓名\_实验报告2; 学号\_姓名\_代码2

### ➤ 提交日期

- 2024年4月21日 22:00前

## 2.1 实验内容2

### ➤ 决策树实现具体任务

### ◆ 基本要求

针对提供的**Titanic数据集**，进行决策树实现，最终需要输出**决策树和分类精度**，并撰写**实验报告3**。

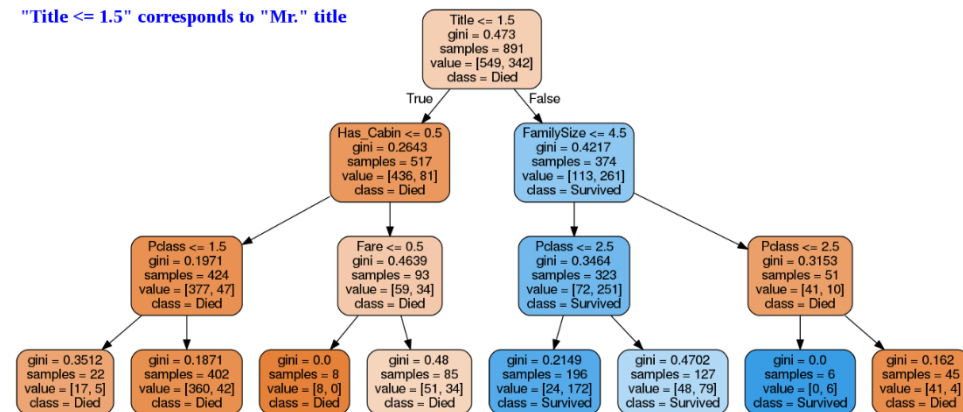
📁 Titanic数据集

<https://www.kaggle.com/c/titanic/data>

### ◆ 加分项

- 可以增加其他数据分析结果
- 可以增加交叉验证等方法寻找最佳tree depth
- ...

"Title <= 1.5" corresponds to "Mr." title



## 2.2 作业提交

### ➤ 提交形式

#### □ 两个压缩包:

- 压缩包1: 实验报告本 (word+pdf)
- 压缩包2: 完整可编译代码 (包括数据集)

#### □ 命名方式: 学号\_姓名\_实验报告3; 学号\_姓名\_代码3

### ➤ 提交日期

- 2024年4月21日 22:00前