

机器学习基础实验

实验工具：scikit-learn, pytorch, matplotlib

实验一：数据分析及可视化

目的：熟悉数据读入及分析的流程，熟悉 scikit-learn

机器学习工具

内容：

- 下载 Iris plants dataset
- 读入数据，并进行可视化，将不同类别的数据机型可视化，并给出结果，对比各个属性

实验二：对比不同线性模型

目的：熟悉最小二乘法，逻辑回归

内容：

- 利用最小二乘法，对不同类别进行回归，并进行可视化，以此作为判断标准，对测试数据进行拟合，计算准确率，并可视化。（训练集，测试集自行划分 5: 5）
- 利用逻辑回归对不同类别进行分类，并计算准确率
- 对比加入 LASSO 以后，对于模型的影响

实验三：决策树及贝叶斯分类器

目的：熟悉和构建基于决策树和贝叶斯分类

- 利用决策树进行数据分类，同时输出决策树构建过程，分析每次为什么选取该特征进行分类
- 构建贝叶斯分类器，并计算准确率，对比贝叶斯分类器和决策树，及线性模型

实验四：神经网络及支持向量机

目的：构建基于 FC 和 CNN 的神经网络，及支持向量机

- 利用 pytorch 构建基于 FC 和 CNN 的神经网络，都仅 1 层
- 构建基于 SVM 的分类器
- 对结果进行对比和可视化，同时对比线性方法

实验五：熟悉聚类方法

目的：熟悉 K-Means，基于密度聚类，和基于层次聚类

- 利用上述三种方法对数据集进行聚类
- 可视化聚类结果，并进行对比