



南京财经大学

《电子商务导论》 课程实验 报告 3

2021 年 5 月 21 日

教师：朱桂祥

学生：小明
学号：212019***

Problem 1

1. 通过 Python 中 Scikit-Learn[1] 载入名为 `load_wine`（红酒）数据集，选用 Scikit-Learn 中任意一种分类器进行训练集和测试集的划分，并利用训练集的样本对分类器进行训练，最终在测试集上进行分类预测，输出模型在测试集上的准确率（代码采用截图方式呈现）。

Solution:

- (1) 载入数据
- (2) 划分训练集和测试集
- (3) 在训练集上训练分类器
- (4) 在测试集上进行分类，计算分类器的准确率

Problem 2

2. 简述 K-Means 算法 [2] 的原理，以及模型的输入、输出及聚类过程（流程）。
Solution:

Problem 3

附加题. 使用 Python 中的 `numpy` 库生成一组随机数据: 样本大小为 100, 特征数为 3, 采用 Scikit-Learn 库中 K-Means 算法进行聚类, 并对聚类结果进行可视化展示。

Solution:

References

- [1] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [2] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern Recognition*, 36(2):451–461, 2003.