03-Machine Learning Fundamentals

课件：数据分析与机器学习\_2022年5月27日.pdf

试验环境链接：https://pan.baidu.com/s/1C43EqMM9oV66GLT-ikkFKg

提取码：n24t

培训录制链接：https://cisco.webex.com/recordingservice/sites/cisco/recording/playback/7594456fbf8a103abf1c00505681061a

密码：59JcU2SD

本文档所在的子目录machine learning下，有二本关于机器学习的书籍。由于机器学习也是当前非常热门的一门学科分支和职业，网上相应的书籍也非常多，可以自行搜索。

请回答下面题目

1. Python语言中关于机器学习涉及到哪些第三方工具库？请列举出来（自行在网上搜索）
2. 机器学习会用到大量数据运算，特别是矩阵运算。请用Python演示一个4x4矩阵的加法，减法，乘法（包括矩阵点乘和矩阵相乘）
3. 在Python代码的调试中，我们经常采用print输出值查看的方法，请问在Python中，print有几种格式化输出？。
4. 请列举出在Python中机器学习的7大步骤。
5. 机器学习是需要从大量的用户数据中找出普通规律，但是数据来源的多样化，用户数据可能存在问题，需要进行预处理，请列出出数据预处理的8大策略。
6. 本培训使用sk-learn机器学习库，在sk-learn中有几种常用的算法？回归类算法要对数据如何处理才能满足回归要求？
7. 针对问题6的各种算法，列出算法结果好坏的评估指标。
8. 什么是归一化？什么是标准化？在机器学习的过程中，它们有什么作用？
9. 请使用提供的数据house.csv，使用回归类算法找出房价和其它自变量之间的关系。（按照7大步骤的顺序），并输出评估指标。
10. 回归类算法的本质就是数据拟合，数据拟合有几种情况，请用图示说明。
11. 请使用sk-learn库提供的boston数据，分析房价和其它自变量之间的关系，并输出评估指标。
12. 重复问题10，如何提高评估指标？并用新的方法重新训练，比较评估指标是否提高？（提示：采用升维的方式，使用多项式拟合）
13. 什么是K-means算法，它属于机器学习中的哪一类？
14. 请使用提供的数据mobile.xls，对用户数据进行分析，并结合可视化方法，判断这些用户生活在什么社区圈？

要求：

1，以mobile.xls为数据来源

2，数据进行归一化或标准化处理

3，先对数据进行聚类

4，对每一类数据进行可视化，给出结论