第6章 使用sklearn构建模型

教案

**课程名称：**Python数据分析与应用

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**64学时（其中理论36学时，实验28学时）

**总学分：**4.0学分

**本章学时**：10学时

# 材料清单

* 1. 《Python数据分析与应用》教材。
  2. 配套PPT。
  3. 数据。
  4. 代码。
  5. 引导性提问。
  6. 探究性问题。
  7. 拓展性问题。

# 教学目标与基本要求

### 教学目标

重点介绍聚类、分类和回归三类sklearn数据分析技术的基本任务对应的数据分析建模方法及实现过程。同时，每一类又有对应的多种评估方法，能够评价所构建模型的性能优劣。通过这一章的学习，基本能够掌握常用的模型构建与评估方法，可在以后的数据分析过程中采用适当的算法并按所介绍的步骤实现综合应用。

### 基本要求

1. 掌握sklearn转换器，评估器的使用。
2. 掌握sklearn数据标准化与数据划分。
3. 掌握sklearn中聚类，分类，回归模型的构建。
4. 掌握sklearn中聚类，分类，回归模型的评价。

# 问题

### 引导性提问

引导性提问需要教师根据教材内容和学生实际水平，提出问题，启发引导学生去解决问题，提问，从而达到理解、掌握知识，发展各种能力和提高思想觉悟的目的。

1. 常见的模型算法使用场景有哪些？
2. 不同的场景之间有什么区别？
3. scikit-learn名字的由来？

### 探究性问题

探究性问题需要教师深入钻研教材的基础上精心设计，提问的角度或者在引导性提问的基础上，从重点、难点问题切入，进行插入式提问。或者是对引导式提问中尚未涉及但在课文中又是重要的问题加以设问。

1. scikit-learn的标准化和第5章的有何不同？
2. 聚类和分类的区别是什么？
3. 回归和分类的区别又是什么？

### 拓展性问题

拓展性问题需要教师深刻理解教材的意义，学生的学习动态后，根据学生学习层次，提出切实可行的关乎实际的可操作问题。亦可以提供拓展资料供学生研习探讨，完成拓展性问题。

1. 不同的算法，可解释性不同，能否挑选一种可解释性强的算法对算法结果进行解释？
2. 除了PCA降维以外，可以使用RandomForest模型进行特征的筛选，该如何做？

# 主要知识点、重点与难点

### 主要知识点

1. sklearn转换器，评估器的使用。
2. sklearn数据标准化与数据划分。
3. sklearn中聚类，分类，回归模型的构建。
4. sklearn中聚类，分类，回归模型的评价。

### 重点

1. sklearn转换器的使用方法。
2. sklearn估计器的使用方法。
3. 聚类模型的构建与评价。
4. 分类模型的构建与评价。
5. 回归模型的构建与评价。

### 难点

1. sklearn转换器的使用。
2. sklearn估计器的使用。
3. 分类模型构建。
4. 回归模型构建。

# 教学过程设计

### 理论教学过程

1. 加载datasets模块中的数据集。
2. 将数据集划分为训练集和测试集。
3. 使用sklearn转换器进行数据预处理与降维。
4. 使用sklearn估计器构建聚类模型。
5. 评价聚类模型。
6. 使用sklearn估计器构建分类模型。
7. 评价分类模型。
8. 使用sklearn估计器构建回归模型。
9. 评价回归模型。

### 实验教学过程

1. 加载datasets模块自带数据集。
2. 划分数据集。
3. 使用sklearn转换器进行数据预处理与降维。
4. 构建与评价聚类模型。
5. 构建与评价分类模型。
6. 构建与评价回归模型。

# 教材与参考资料

### 教材

黄红梅，张良均．Python数据分析与应用[M]．北京：人民邮电出版社．2018．

### 参考资料

[1] 张良均．Python数据分析与挖掘实战[M]．北京：机械工业出版社．2015．

[2] 张良均．Python与数据挖掘[M]．北京：机械工业出版社．2016．