

实战案例：Python机器学习小案例源码 -- 骨科疾病预测

作者： [Robin](#)

日期： 2018/02

提问： [小象问答](#)

数据集来源： [kaggle](#)

1. 案例描述

近年来，人工智能（AI）发展迅速，从AlphaGo连败人类棋手，到商场里随处可见的智能机器人，人工智能已经从实验室走向了大众，不论是舆论关注度还是相关领域的投资，都在节节增长。更重要的是，人工智能技术也达到了新的阶段，在工业界、医疗、SaaS、农业等等各行各业的应用都引起了巨大的势能。这其中，应用增长率最高的当属AI在医疗领域的应用。

该案例通过数据分析的方法探索骨科就诊人员的数据，建立一个简单的机器学习模型，用于预测就诊人员是否患有骨科疾病。该案例适合初次接触数据分析、机器学习及人工智能的读者。

2. 数据集描述

- 该数据集由Kaggle[提供](<https://www.kaggle.com/uciml/biomechanical-features-of-orthopedic-patients>)
- 数据字典
 - **pelvic_incidence**: 骨盆入射角，浮点型
 - **pelvic_tilt numeric**: 骨盆倾斜，浮点型
 - **lumbar_lordosis_angle**: 腰椎前凸角度，浮点型
 - **sacral_slope**: 骶骨倾斜角，浮点型
 - **pelvic_radius**: 盆腔半径，浮点型
 - **degree_spondylolisthesis**: 腰椎滑脱程度，浮点型
 - **class**: 病人是否患病，字符型：Abnormal, Normal

3. 任务描述

- 根据病人的6项医疗数据，推断该病人是否患有骨科疾病

4. 主要代码解释

- 代码结构

```
├─ data.csv          # 数据文件
├─ main.ipynb        # jupyter notebook演示文档
├─ model.pkl         # 保存的训练好的模型（需要运行程序才能得到）
└─ proj_readme.pdf   # 案例讲解文档
```

- 具体代码请参照main.ipynb

5. 案例总结

- 该项目通过学习kNN模型，基本能“准确”地预测出病人是否患有骨科疾病，同时也包括了以下概念：
 - 数据处理
 - 数据分析和机器学习的基本步骤
 - 数据可视化

6. 课后练习

- 熟悉Python的读者，可以试着将以上代码写成.py文件
- 试着只使用6个特征中的一些，观察对结果的影响；
- 考虑只用准确率能否真实地体现模型的好坏，是否有其他的评价指标？

参考资料

1. [10分钟走入Pandas](#)
2. [matplotlib教程](#)
3. [seaborn教程](#)
4. [scikit-learn教程](#)