

分析流程



分析步骤

1. 通过训练集数据来建立XGBoost分类模型。
2. 通过建立的XGBoost来计算特征重要性。
3. 将建立的XGBoost分类模型应用到训练、测试数据, 得到模型的分类评估结果。
4. 由于XGBoost具有随机性, 每次运算的结果不一样, 若保存本次训练模型, 后续可以直接上传数据代入到本次训练模型进行计算分类。
5. 注: XGBoost无法像传统模型一样得到确定的方程, 通常通过测试数据分类效果来对模型进行评价。

详细结论

输出结果1: 模型参数

[复制](#)

| 参数名 | 参数值 |
|--------------|--------|
| 训练用时 | 1.457s |
| 数据切分 | 1 |
| 数据洗牌 | 是 |
| 交叉验证 | 10 |
| 基学习器 | gbtree |
| 基学习器数量 | 100 |
| 学习率 | 0.1 |
| L1正则项 | 0 |
| L2正则项 | 1 |
| 样本征采样率 | 1 |
| 树特征采样率 | 1 |
| 节点特征采样率 | 1 |
| 叶子节点中样本的最小权重 | 0 |

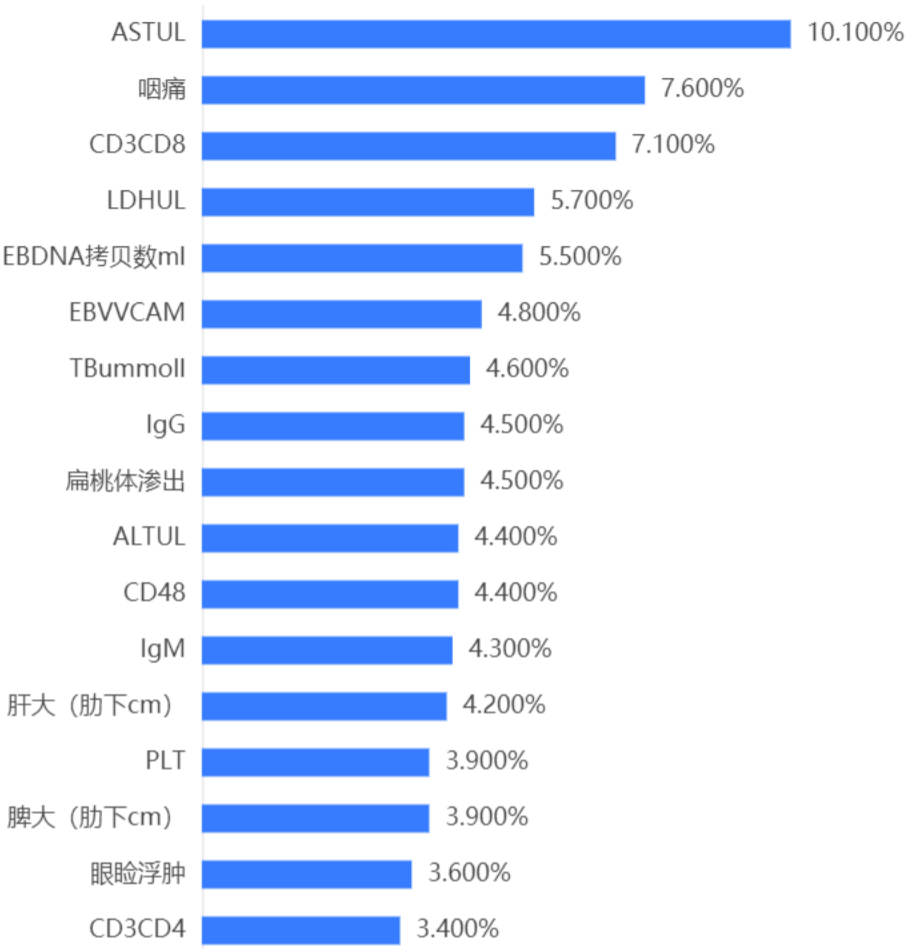
| | |
|--------|----|
| 树的最大深度 | 10 |
|--------|----|

图表说明：

上表展示了模型各项参数配置以及模型训练时长。

输出结果2：特征重要性

柱形图

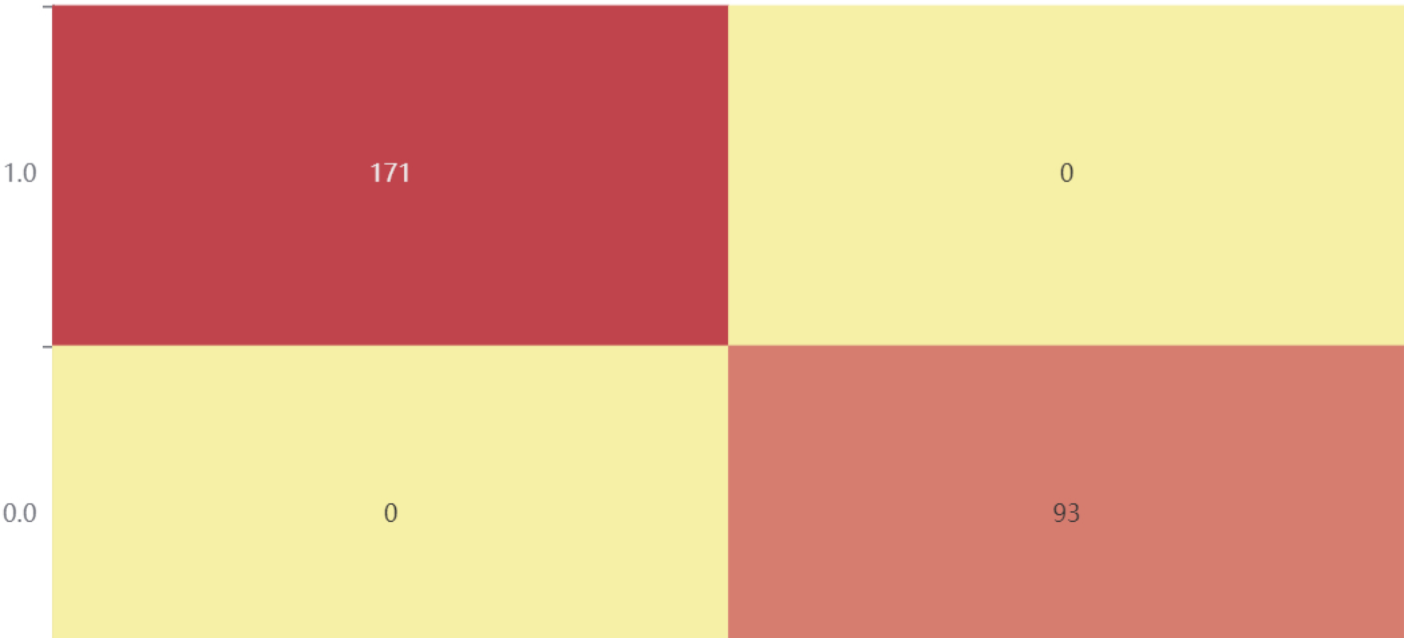


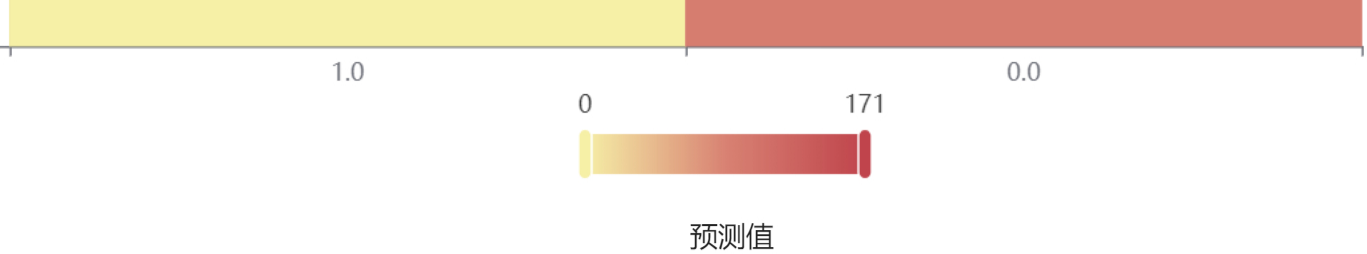
图表说明：

上柱形图或表格展示了各特征（自变量）的重要性比例。

输出结果3：混淆矩阵热力图

训练数据





图表说明：
上表以热力图的形式展示了混淆矩阵。

输出结果4：模型评估结果

复制

| | 准确率 | 召回率 | 精确率 | F1 |
|-------|-------|-------|-------|-------|
| 训练集 | 1 | 1 | 1 | 1 |
| 交叉验证集 | 0.647 | 0.647 | 0.703 | 0.641 |

图表说明：
上表中展示了交叉验证集、训练集和测试集的预测评价指标，通过量化指标来衡量XGBoost的预测效果。其中，通过交叉验证集的评价指标可以不断调整超参数，以得到可靠稳定的模型。

- 准确率：预测正确样本占总样本的比例，准确率越大越好。
- 召回率：实际为正样本的结果中，预测为正样本的比例，召回率越大越好。
- 精确率：预测出来为正样本的结果中，实际为正样本的比例，精确率越大越好。
- F1：精确率和召回率的调和平均，精确率和召回率是互相影响的，虽然两者都高是一种期望的理想情况，然而实际中常常是精确率高、召回率低，或者召回率低、但精确率高。若需要兼顾两者，那么就可以用F1指标。

输出结果5：模型预测与应用

请选择文件所在路径

模型预测 ?

☐ 数据是否包括实际因变量值Y

图表说明：

- 系统会自动保存模型，需要注意的是：在机器学习中的XGBoost算法保存的模型是非常复杂的，不是类似于线性回归那样可以用一个公式保存，系统以二进制文件方式进行序列化保存。
- 由于XGBoost具有随机性，每次训练的模型可能不一致，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算预测。
- 若删除本分析报告将会直接删除模型的缓存。

参考文献

[1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from <https://www.spsspro.com>.

[2] Chen T , Guestrin C . XGBoost: A Scalable Tree Boosting System[J]. ACM, 2016.