实验六 sklearn模块使用

1. 实验目的

学会sklearn基本函数的使用

1. 实验内容

这个数据集是关于一家葡萄牙银行机构从2008年到2010年的直接电话营销活动，旨在促进现有客户的定期存款，总共4万多条记录，每一条有21个属性。主要的任务目标是分类预测，用户是否会接受营销。

变量介绍

银行客户信息：

1.age 年龄（数字）

2.job 工作类型

3.marital 婚姻状况

4.education 受教育程度

5.default 是否有信用违约

6.housing 是否有住房贷款

7.loan 是否有个人贷款

与联络相关信息：

8.contact 联系类型

9.month 年度最后一次联系的月份

10.date\_of\_week 最后一次联系的星期

11.duration 上一次联系的通话时长

其他属性：

12.campaign 在本次营销周期内与该客户的总通话次数

13.pdays 上次营销到现在已经过了多少天

14.previous 在本次营销之前和客户联系过几次

15.poutcome 上一次营销活动的结果

社会和经济相关属性：

16.emp.var.rate 就业变动率-季度指标

17.cons.price.idx 消费物价指数-月度指标

18.cons.conf.idx 消费者信心指数-月度指标

19.euribor3m 欧元同业拆借利率3个月-每日指标

20.nr.employed 员工数据-季度指标

输出变量（目标）：

21.y 是否认购定期存款

1载入数据及数据观察

使用info,describe函数

2.数据处理

2.1直接删除job和matrial中unknow记录

2.2变量education,default,housing和loan缺失数过多，不宜直接删除，采用预测方法拟合缺省值。

2.3.分类变量数值化处理

2.4.数据规范化

3.模型的训练与评估

3.1数据重采样

3.2划分数据集

3.3使用逻辑回归、随机森林、支持向量机进行分类

3.4模型评估

1. 实验要求

1.提交实验报告（pdf格式）

2.在实验报告运行材料中提交含有桌面图标的运行截图