基于心血管疾病的多标签分类

1. 数据集统计

数据处理与心衰数据处理类似，不同之处在于 ：

* 未加对照组，仅提取出了要研究的几种心血管疾病数据集。
* 数据缺失值处理上对缺失值进行了0值填充。
* 目前使用的是所有特征， 未进行降维，数目为270多个

|  |  |  |
| --- | --- | --- |
| 疾病名称 | 高血压组数据量 | 糖尿病组数据量 |
| 脑卒中 | 5956 | 1933 |
| 心衰 | 779 | 300 |
| 急性心梗 | 740 | 290 |
| 心源性猝死 | 130 | 42 |
| 心肌缺血 | 857 | 280 |
| 肺心病 | 87 | 29 |
| 心脏功能病变 | 2462 | 650 |
| 冠心病 | 19761 | 7000 |

1. 问题转换：二分类预测

选择大约80%作为训练集，20%作为测试集。

高血压

逻辑斯特回归模型结果

最大最小归一化，过采样对照组数据集后

|  |  |  |
| --- | --- | --- |
| 名称 | ROC | 训练数据集大小 |
| 脑卒中 | 0. 688971929108 | (exp:4761 ,con:5220 ) |
| 心衰 | 0. 678974234199 | exp:608 ,con:1010 |
| 急性心梗 | 0. 637900509027 | exp:583 ,con:698 |
| 心源性猝死 | 0. 618636733057 | exp:103 ,con:284 |
| 心肌缺血 | 0. 57900430815 | exp:682 ,con:1137 |
| 肺心病 | 0. 633996952514 | exp:68 ,con:309 |
| 心脏功能病变 | 0. 607603166427 | e xp:1968 ,con:2220 |
| 冠心病 | 0.676451652362 | exp:15836 ,con:16123 |

SVM预测结果

|  |  |  |
| --- | --- | --- |
| 名称 | ROC | 训练数据集大小 |
| 脑卒中 | 0. 707931332331 | (exp:4761 ,con:5220 ) |
| 心衰 | 0. 718223872612 | exp:608 ,con:1010 |
| 急性心梗 | 0. 693227546448 | exp:583 ,con:698 |
| 心源性猝死 | 0. 784992395624 | exp:103 ,con:284 |
| 心肌缺血 | 0. 690541312254 | exp:682 ,con:1137 |
| 肺心病 | 0. 884186225471 | exp:68 ,con:309 |
| 心脏功能病变 | 0. 629307045327 | e xp:1968 ,con:2220 |
| 冠心病 | 0. 733479217558 | exp:15836 ,con:16123 |

糖尿病

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | ROC | | |
| 脑卒中 | | | 0. 646025385382 |
| 心衰 | | 0. 768244726253 | |
| 急性心梗 | | 0. 368398218314 | |
| 心源性猝死 | | 0. 777898694104 | |
| 心肌缺血 | | 0. 680580273686 | |
| 肺心病 | | 0. 943534726143 | |
| 心脏功能病变 | | 0. 665115171814 | |
| 冠心病 | | 0. 646160853814 | |

备注：

1. 排除严重脑神经疾病、意外死亡及后遗症的疾病
2. 加入风湿性心脏病作为标签
3. 去掉严重疾病2的ICD数据（冠心病）
4. 新类标见

|  |
| --- |
| 心衰人群 |
| 严重肾病人群 |
| 急性心梗/急性冠脉综合症 |
| 脑卒中 |
| 心源性猝死 |
| 肺栓塞 |