逐步代入降维

作为随机森林封装法的对照实验，之前的过滤法得到了28个变量，按照其重要性依次代入神经网络、逻辑回归与支持向量机模型中，经过十折交叉验证查看其AUROC值的变化，如下图所示，28个变量全部使用准确率最高。与之前使用随机森林函数封装法对比，封装法当选择20个变量时（随机森林封装法每次运行结果略有差别，在进行与其他方法对比时我选用最少的特征变量即20个），精确度就接近54个变量全部使用的水平了，对于神经网络算法甚至有所提高，而依次带入模型中计算时，对于逻辑回归使用变量25个准确度才能超过81.99%（随机森林封装法的20个变量），而支持向量机需使用27个变量才能超过81.89%（随机森林封装法的20个变量），而且该方法耗时较长。故考虑快速的得到较少的变量而精确度较高的情况，可以使用随机森林函数过滤并封装来降维。



下表MLP是使用54个变量的神经网络十折交叉验证结果，RF过滤封装降维法是过滤与封装均使用随机森林来降维（使用20个特征值），RF+MLP降维法是过滤过程使用随机森林，之后使用神经网络算法查看准确率来选择特征个数（如上图所示，使用28个特征值准确率最高）。其他同上，LR代表逻辑回归，SVM代表支持向量机。

|  |  |
| --- | --- |
|  | AUROC |
| MLP | 86.20% |
| **RF过滤封装降维法** | 86.31% |
| RF+MLP降维法 | 86.27% |
| LR | 83.16% |
| **RF过滤封装降维法** | 81.99% |
| RF+LR降维法 | 82.45% |
| SVM | 83.00% |
| **SVM过滤封装降维法** | 81.89% |
| RF+SVM降维法 | 82.34% |