

Ex5 record and report

王敏行 id: 2018012386 wangmx18@mails.tsinghua.edu.cn

Part 11: Bayesian network

这里使用 `sklearn.naive_bayes.GaussianNB` 函数，构建朴素贝叶斯网络，对 icu 的生存情况进行预测。

首先对每一个 feature 的分布情况进行检测，借助 K-S test，发现没有 feature 是正态分布的。还尝试了指数分布、指数正态、对数正态，在 $p=0.05$ 的标准下，都不符合。但是可以看出，`elective_surgery` 等部分变量是二值分布的。

虽然 `GaussianNB` 要求样本是正态分布的，但是这里强行用高斯朴素贝叶斯模型进行训练和预测，仍然有不错的表现，如下所示：

```
sklearn-Gaussian Naive Bayes  
train acc:0.75580 validation acc:0.76299  
train err:0.24420 validation err:0.23701
```

准确率和 Ex3 中的 SVM 相比接近，且不存在过拟合的现象。虽然样本的分布并不符合正态分布的要求，但是模型的好表现说明 `GaussianNB` 的鲁棒性很高。