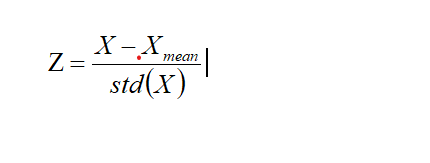
阈值来源于特征标准化公式，由于表中数据不均衡，得到表中违约率为：0.1727%说明违约的数据量远远小于正常的数据量，需要进行样本均衡化;Amount变量，范围较大，远远大于V1-V28，需要进行标准化处理。Time变量无实际意义，丢弃即可，所以使用特征标准化公式计算阈值。



Z为标准化后的数据，X为原始数据，Xmean为原始数据的均值，std(X)为原始数据的标准差。Python中提供了StandardScaler模块可以进行特征标准化。

得到最终阈值为0.05