朴素贝叶斯的实现

姓名：储醉

学号：MG1433009

邮箱: chuqi77@gmail.com

联系方式: 13770317817

(南京大学 计算机科学与技术系, 南京 210093)

# 实现细节

字典构造：将所有的文本分词后汇总，并将stopwords中的词去掉，得到的字典总条数为28266。

TF 提取：每个帖子生成一条长为28266的向量，每位上的数字为字典中对应单词在该帖子中出现的次数。

TF-IDF提取：TF值乘以IDF，IDF为log((N+1) / (d+1))，其中N为总贴数，d为有该单词出现的贴数。

NBD：Multinomial naive Bayes的实现。

NBCG：Gaussian naive Bayes的实现

NBCD：最简单的等区间分段法，当赋予落在最后的区间内的点以适当的较大数值时，正确率有少许提升。

# 结果

使用TF特征的NBD：mean为0.75，std为0.03

使用TF-IDF特征的NBCG：mean为0.77，std为0.03

使用TF-IDF特征的NBCD：mean为0.84，std为0.04