



谷歌也是这样操纵，有一堆网页，有一堆腾讯新闻，有向图指向另一个网页





无监督算法，用来挖掘主题





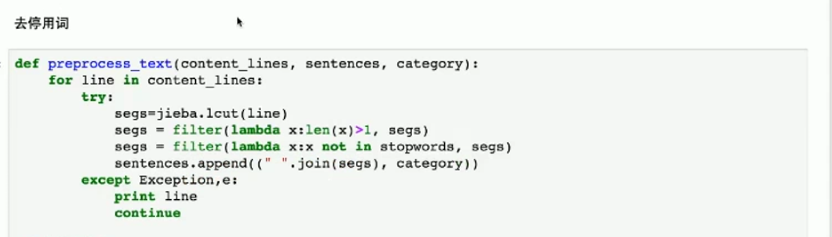


给这些词标号

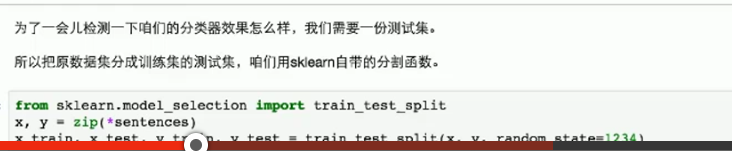
朴素贝叶斯





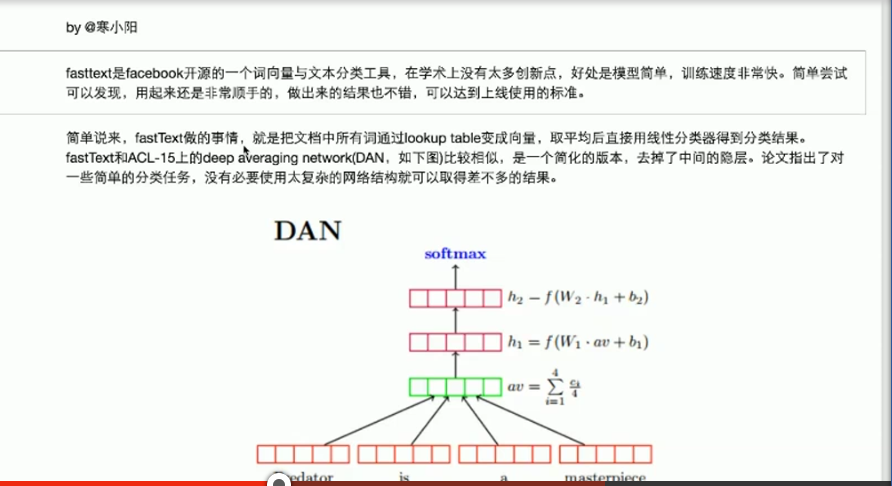






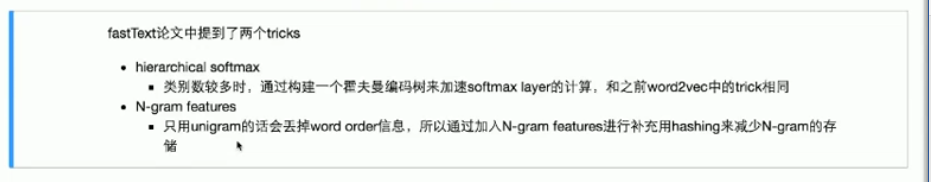






线性分类器

把这些词抽取出来的，把每个词做处理，丢到分类词



优化并不加快准确率，是速度

Fasttext 用的python接口

神经网络