案例：文本分词

对于段落，可以按照语句拆分，也按照单词拆分。

文本向量化处理，词袋模型，对词袋模型归一化处理。

在训练语义分析模型时，需要对每一个段落的每一个样本做向量化处理，以样本的语义(标签)建立一个训练样本集。

比如样本的输入x：这个酒店的卫生很好！

样本的输出y：好评

此时需要一种算法可以把一句话转化成一个特征向量，需要改向量用数字来表达语义。

一句话的语义很大程度上取决于某个单词出现的次数，所以可以把语句中所有出现的单词作为特征名，以单词出现的次数来构建模型，称为词袋模型。

doc = “This hotel is very bad very bad. The toilet in this hotel smells bad. The environment of this hotel is very good.”