训练集:验证集:测试集=6:2:2

一般的时候, 我们组织训练数据是以文件夹的形势组织, 里面每一个文档是一个训练文档, 结构如下：

data

class1

doc1.txt

doc2.txt

doc3.txt

class2

doc1.txt

doc2.txt

doc3.txt

现在需要将每一个分类的文档分为三个部分: 训练数据, 验证数据,测试数据