文本表示及特征选择

常用的特征有：TF/IDF，互信息量，信息增益，X2统计

特征选择：选择具有代表性的特征，

特征提取：降维