**1、项目背景：**

海伦女士一直使用在线约会网站寻找适合自己的约会对象。尽管约会网站会推荐不同的人选，但她并不是喜欢每一个人。经过一番总结，她发现自己交往过的人可以进行如下分类：

* 不喜欢的人
* 魅力一般的人
* 极具魅力的人

**2、项目数据**

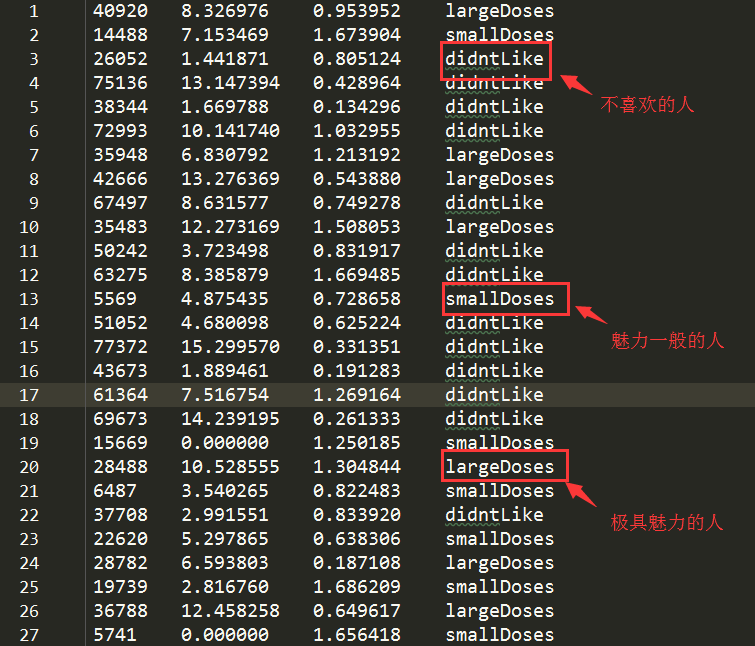
 海伦收集约会数据已经有了一段时间，她把这些数据存放在文本文件datingTestSet.txt中，每个样本数据占据一行，总共有1000行。

[datingTestSet.txt数据集下载](https://github.com/Asia-Lee/KNN/blob/master/datingTestSet.txt)

 海伦收集的样本数据主要包含以下3种特征：

* 每年获得的飞行常客里程数
* 玩视频游戏所消耗时间百分比
* 每周消费的冰淇淋公升数

数据格式如下：



**3、K-近邻算法的一般流程**

（1）收集数据：提供文本文件。

（2）准备数据：使用Python解析文本文件。

（3）分析数据：使用Matplotlib画二维扩散图。

（4）测试算法：使用文本文件的部分数据作为测试样本，计算错误率。

（5）使用算法：错误率在可接受范围内，就可以运行k-近邻算法进行分类。