1. 实验题目

特征工程

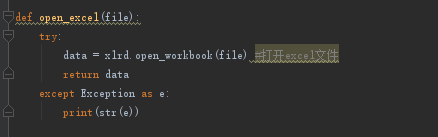
1. 实验目的

将已有的excel表中的数据进行一个特征处理，将表中的数据转换为更好的代表预测模型的潜在问题的特征的过程，从而提高了对未知数据的模型准确性。

1. 实验内容具体操作解释

from sklearn.model\_selection import train\_test\_split

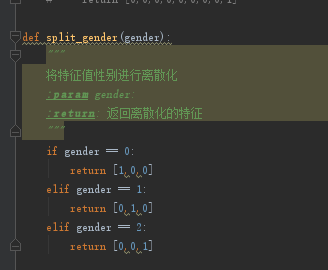
可知用sklearn将 model\_selection模型中选择导入到train\_text\_spit。进入model\_selection数据库筛选出需要导入的数据，在开始特征处理前的预处理操作



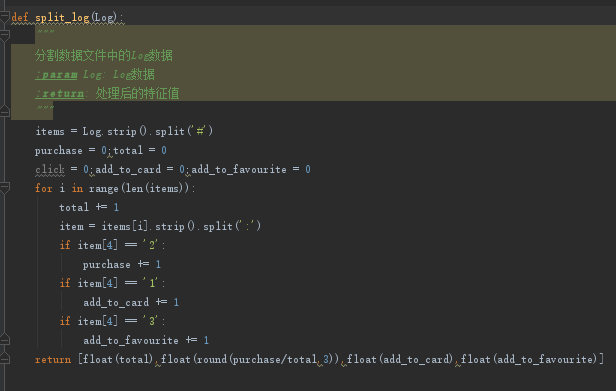
定义一个打开文件的函数，并判断是否有其他错误的情况，以便后续操作



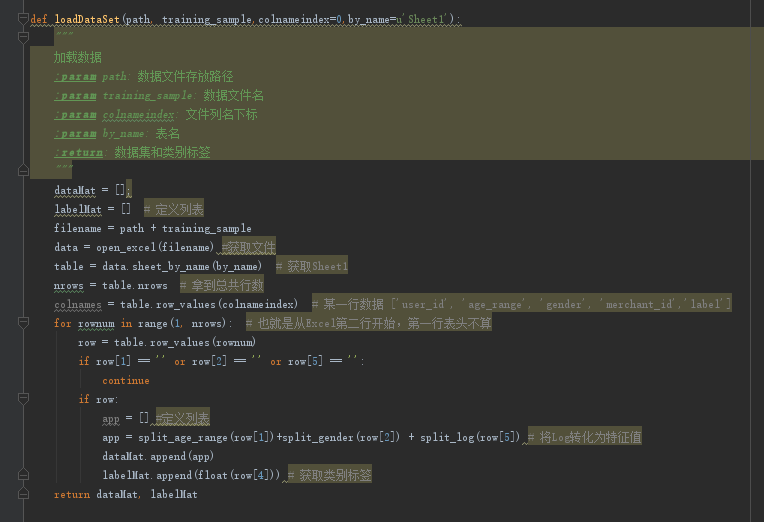
定义一个展开函数，将不同年龄区间的人分在数组不同的下标上



有些数据会最为数组下标来用，若范围过大，无法开出足够大的数组，而恰好实际用到的下标又不多，即直接开数组会造成大量浪费，此时可以离散化，并返回离散化的特征



将leg中的数据进行分割，以便后续处理



读取excel表中的所有数据信息，并调用函数将其转换为特征值