**疾病推断**

本模型使用多项式朴素贝叶斯来预测患者的疾病编码。

基本思路：

将一个患者做过的项目列表转换成一个向量，每个元素反映这个患者在对应项目下所花费的具体金额。然后将这个数值向量进行离散化，根据每个项目占该患者总费用的百分比，将项目分为数值为0-5的6类。

综上，模型的输入为一个数值为0-5，长度为所涉及项目长度的向量。向量每一个元素对应于一个具体的项目费用在整体费用中的权重。

模型根据多项式朴素贝叶斯模型，根据历史数据的训练，获取先验信息。当需要预测时，输入患者的如上所述的向量，输出该患者对应于每一个icd-10编码的概率。

模型当前准确率：

top 1 47.5%

top 10 87.1%