

ML Homework

汪珉凯 3220100975

1

10交叉验证法指的是将样本划分为10个同分布但是不相交的子集，其中1个作为测试集，剩余的9个作为训练集，进行训练和验证。然后更换测试集，重复一共10次训练验证的过程。因此题目中的模型在验证时，由于训练集的正反例数量相同，而模型在这种情况下对于验证集给出的内容会进行随机猜测，因此正确的概率是0.5。留一法指的是仅留下一个样本作为测试集，剩余的99个样本都作为训练集。假设作为测试集的样本是反例，那么训练集里有50个正例、49个反例，正例数量更多，因此我们的模型会将测试集预测为正例，因此正确率为0。如果测试集是正例也同理。所以该模型一定预测错误。

2

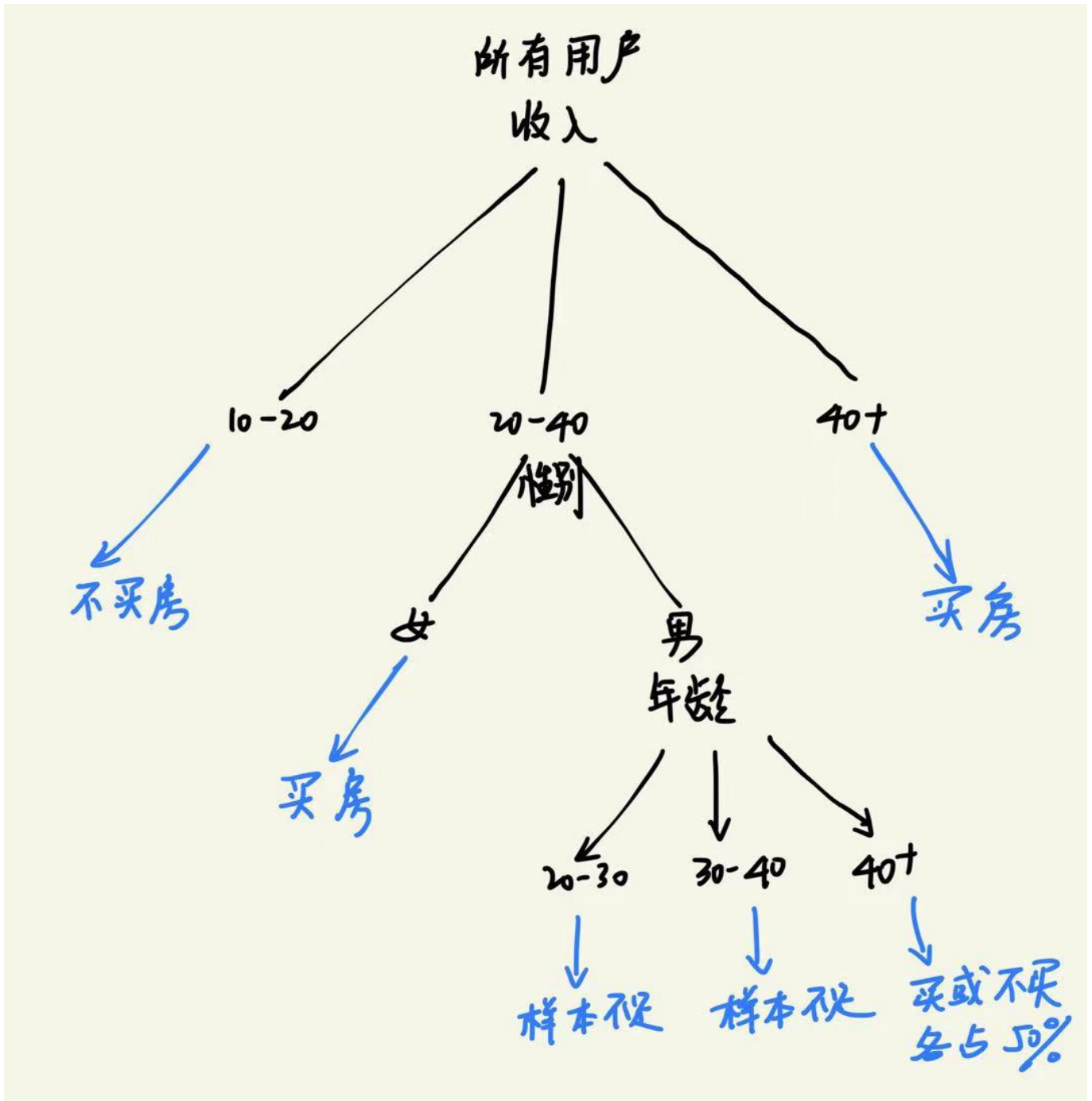
C~1~ = 000000000 ;
C~2~ = 000000111 ;
C~3~ = 000111000 ;
C~4~ = 111000000 ;
海明距离和为27.

3

混淆矩阵如下： $\begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}$

准确率 Accuracy=0.7
精确率 Precision=0.8 召回率 Recall=0.6666
F1 Score = 0.6666

4



5

1. 错误。高偏差一般意味着模型太过简单导致欠拟合，单纯增加样本数量并不能改善这一情况。
2. 错误。有时候我们更喜欢低方差的模型，有时候我们要选择低偏差的模型，不能一概而论。
3. 正确。越复杂的模型越容易过拟合，方差越高。
4. 错误。引入正则项并不一定能解决问题。
5. 错误。如果正则化因子 λ 的值太大，可能会使模型过于简单，损害其性能。