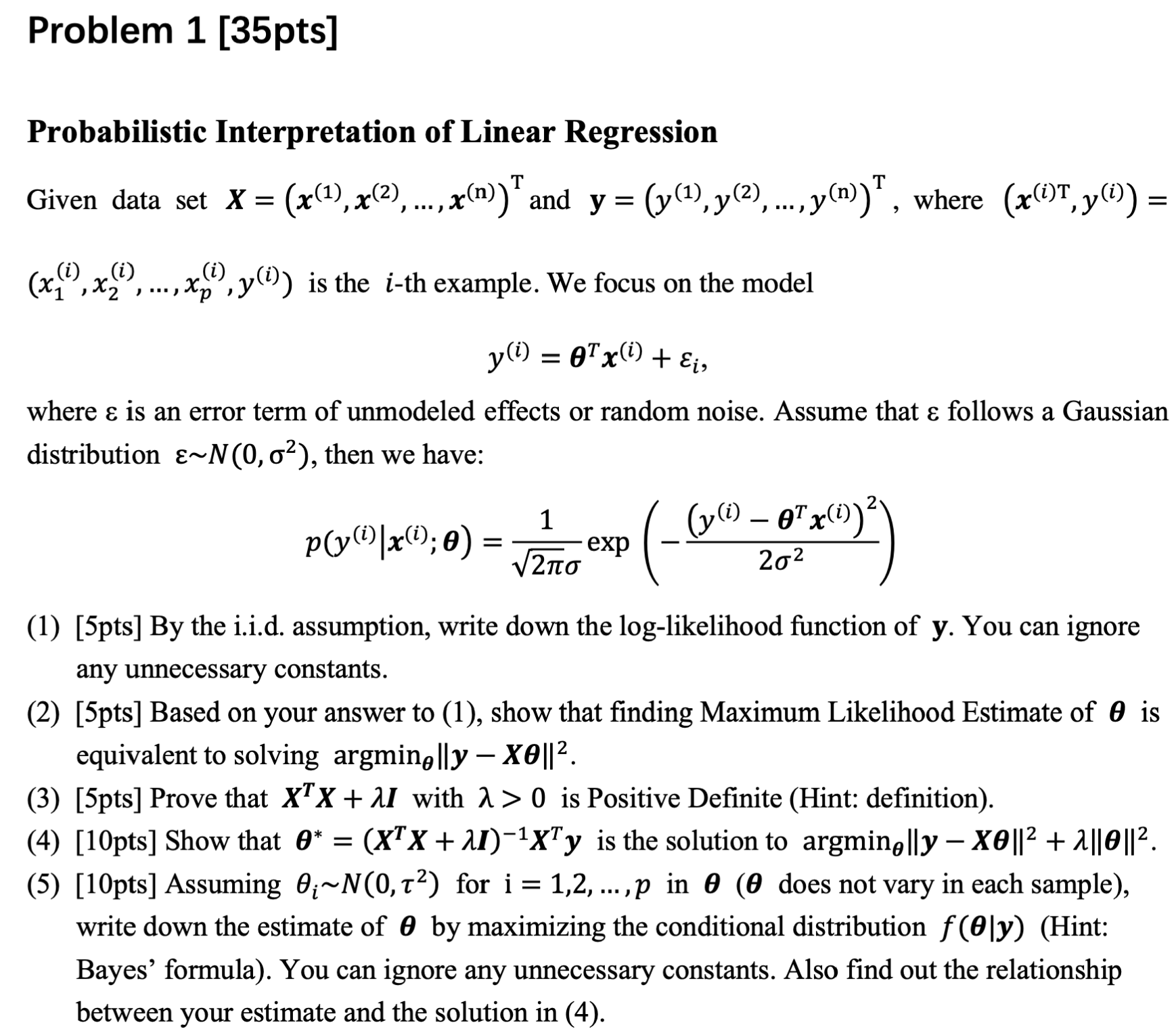
**Assignment 2 作业详解**



1. 写出log likelihood，忽略无用的常数，例如

最终答案为：

也可以进一步的省去，但是负号要保留。

1. 最大化似然估计：
2. 证明一个矩阵 是positive definite的，即对于任意非零的向量，满足： 。
3. 证明是解，则需要证明损失函数的梯度在时为0。

，代入验算，可得。

1. 使用贝叶斯定理：

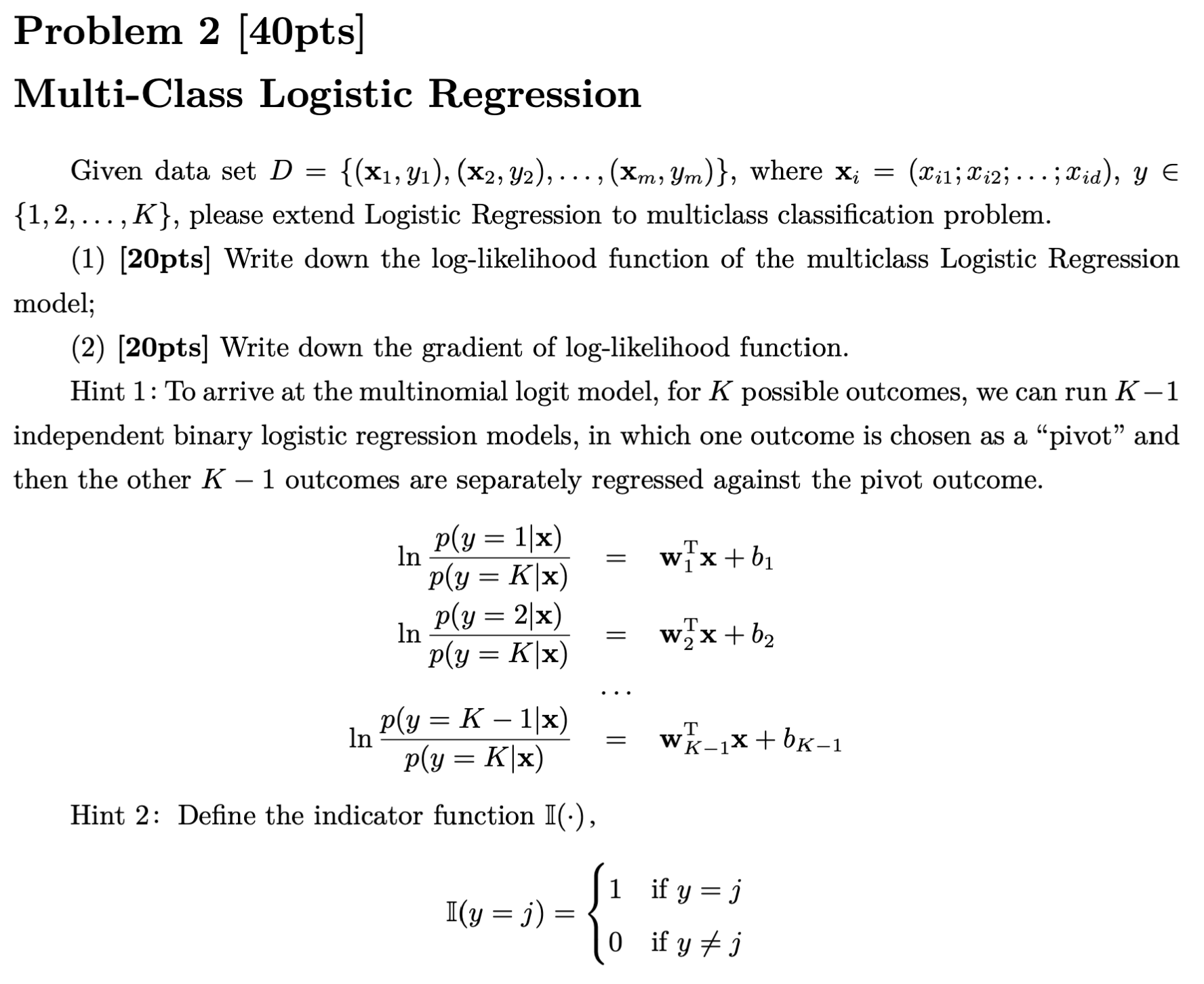
是已知的，可以看做常数，因此：

再代入高斯表达式，以及利用条件可得：

因此，

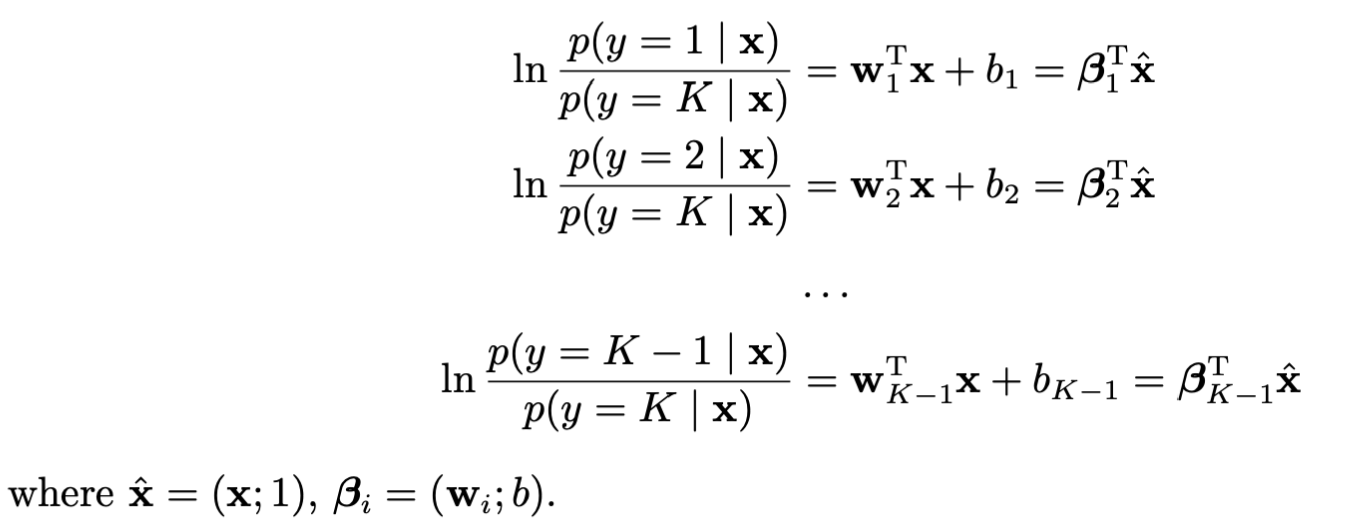
关于求导可得：

与(4)的关联在于。



1. 关键是写出。

回顾一下2 class的情况，是建模：，这里的即选中的pivot。同理，这边把当做pivot。



注意，这里的需要给矩阵加一列常数1作为intercept，对应的给权重矩阵加一列表示bias。

转换： for ，那么我们有：

根据性质，我们把得到的全部加起来（包括）：

那么，可以解得

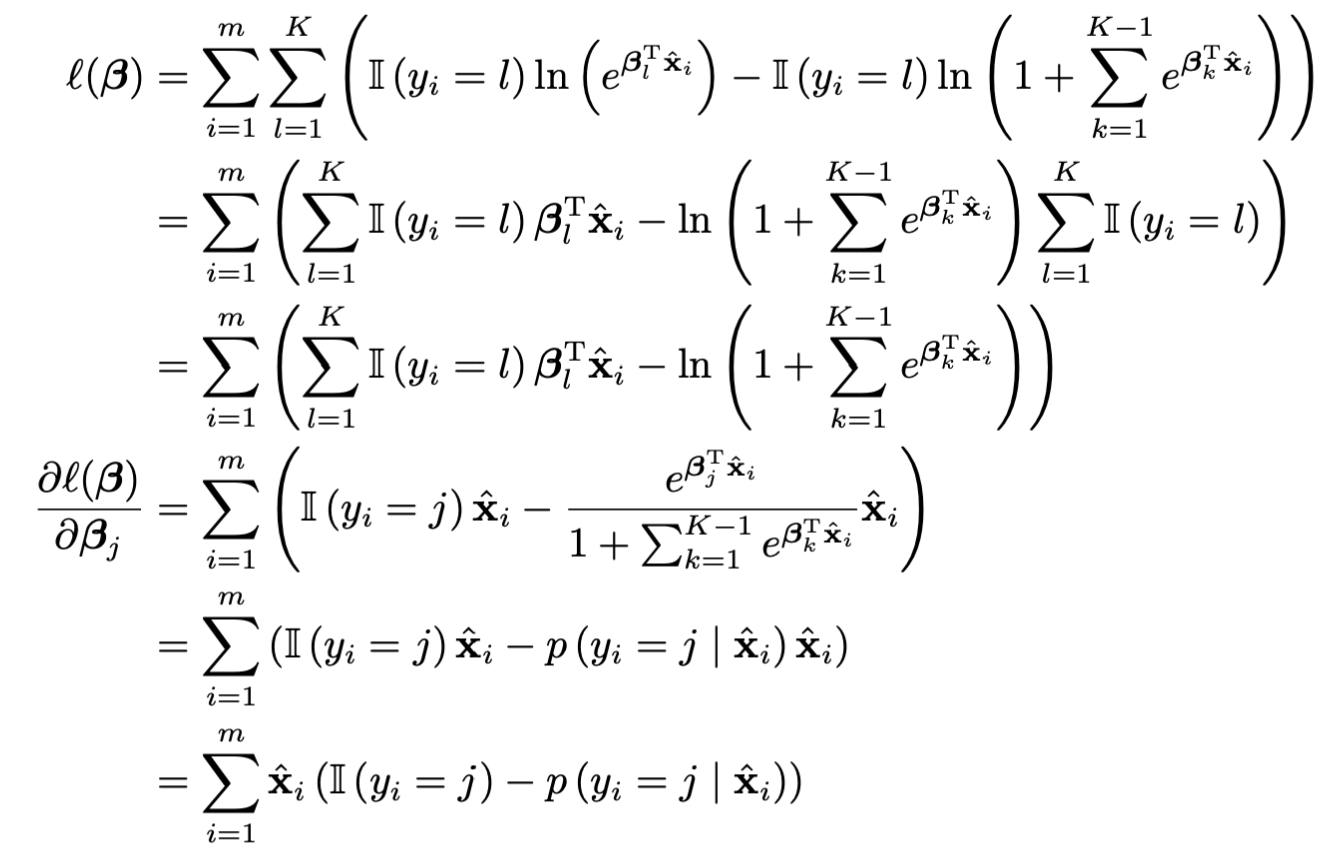
最终我们有：

其中（设），

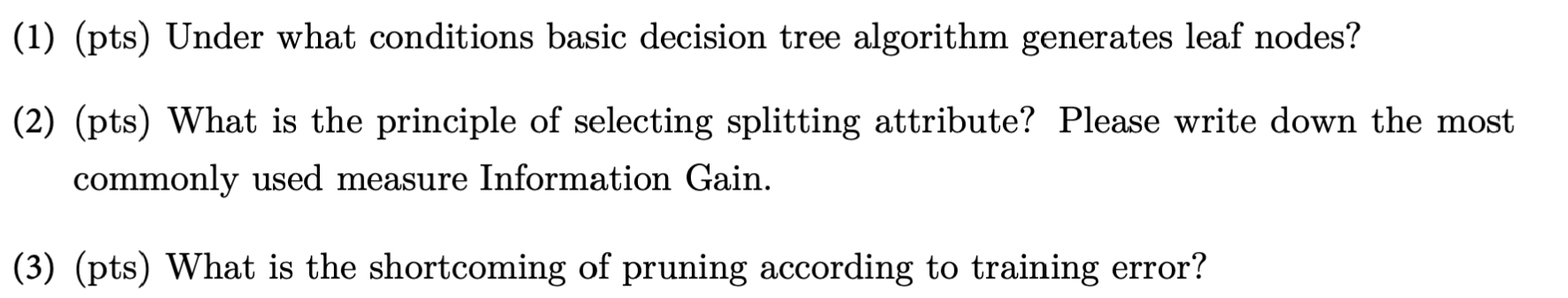
将关于的表达式代入

可得：

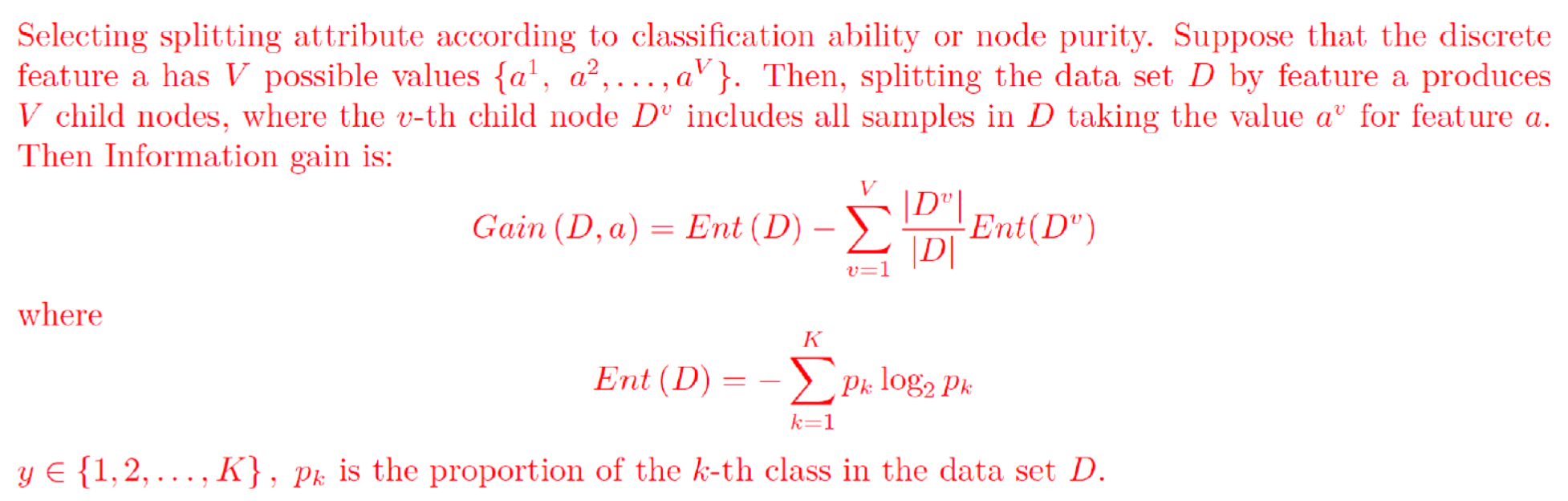
继续化简：



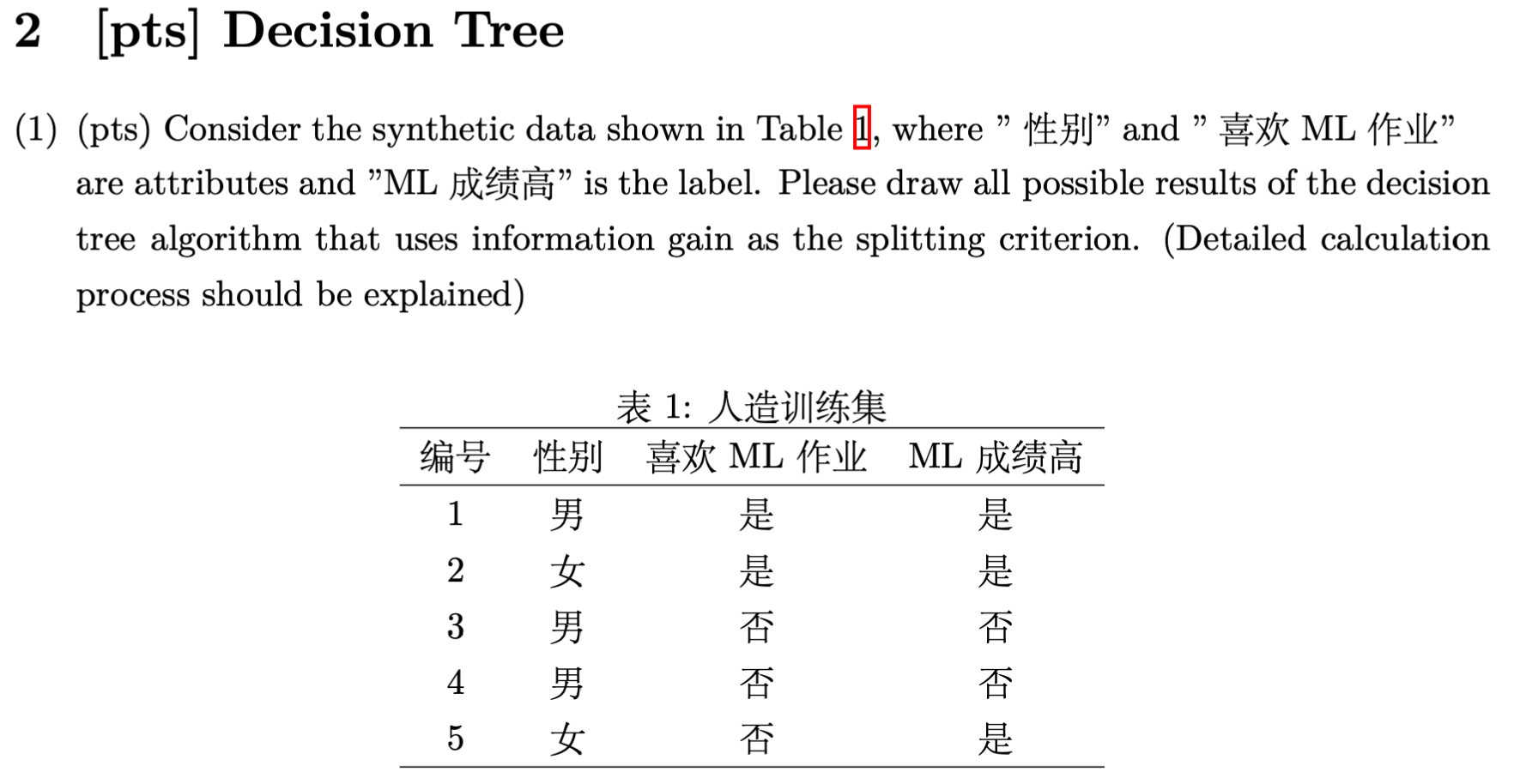
**Lecture 4 Decision Trees**

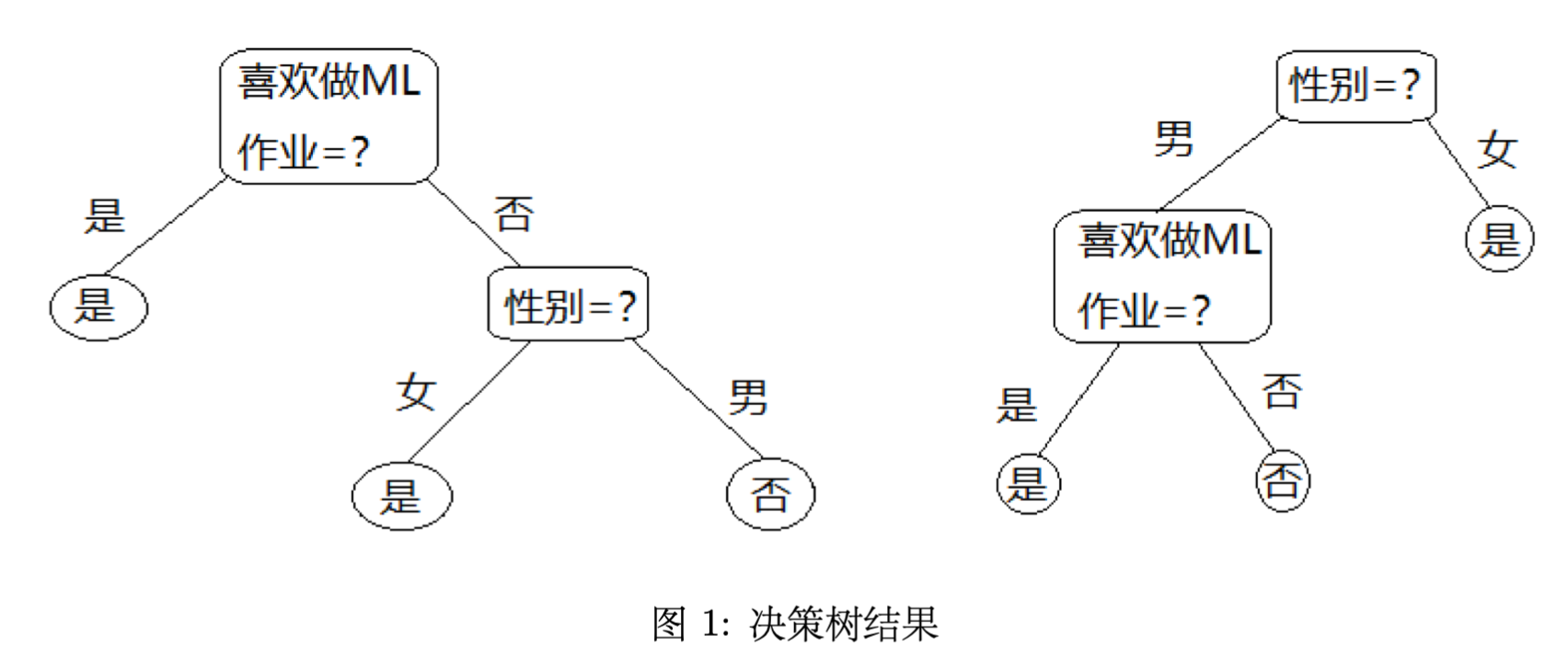


1. 问何时形成树叶节点。简单而言，就是无法继续分裂。通常三种情况：
2. 所有节点属于同一个标签（类）；
3. 所有节点的特征值相同或者特征值集合为空；
4. 没有数据点了
5. 满足一些早停条件（例如树的深度）
6. 最大化信息增量（information gain）。信息增量是通过母节点的信息熵减去分裂后子节点的加权信息熵之和来计算的。



1. 问的是根据training loss剪枝的风险。风险是过拟合。





总共就两个特征，因此按照特征的先后决策树的结果无非就两种。

初始的信息熵：

若是首先按照“喜欢ML作业”生长：

（是）

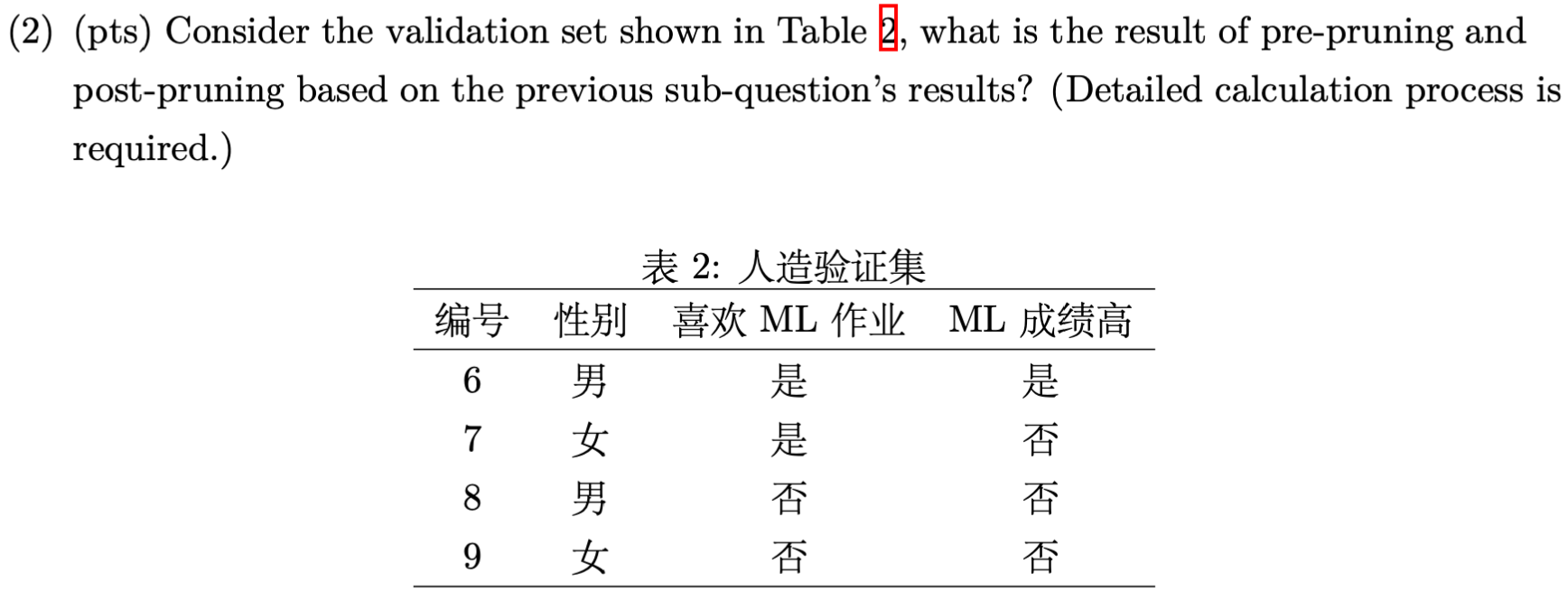
（否）

若是首先按照“性别”生长：

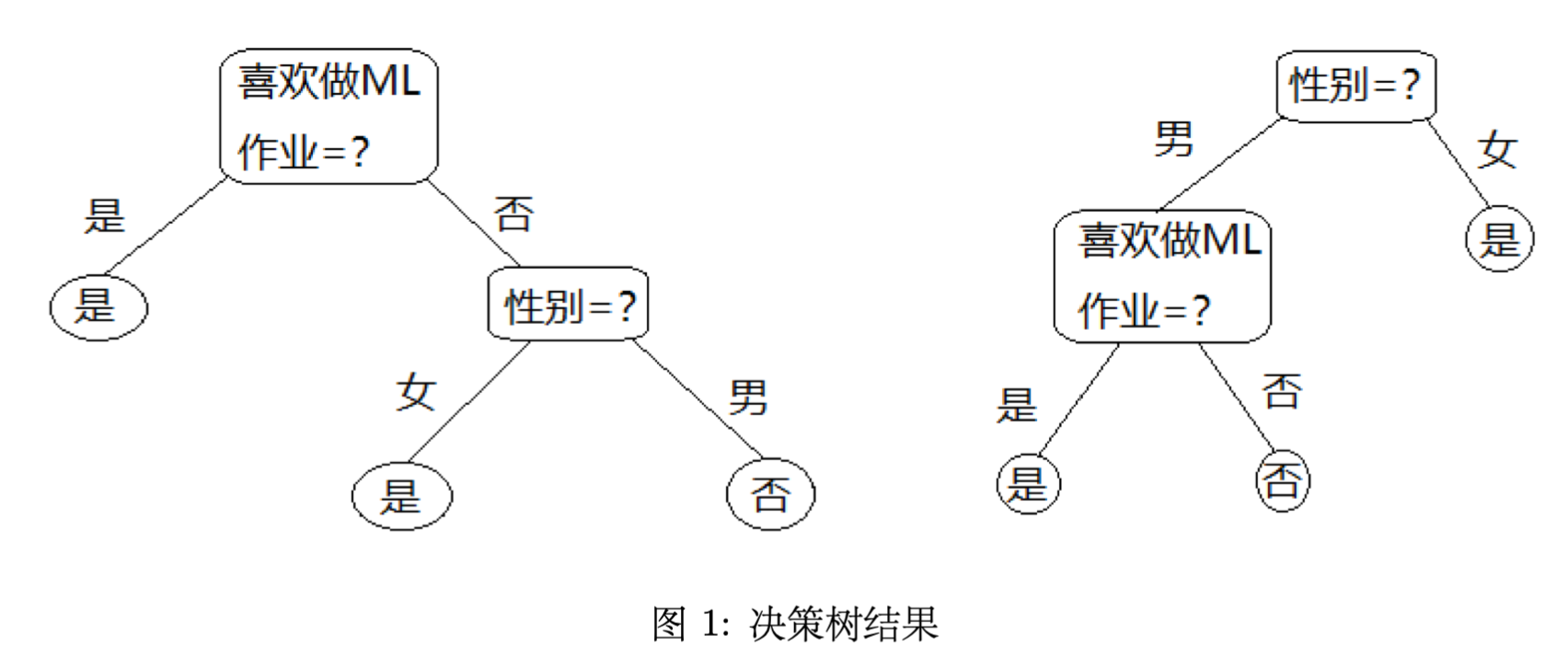
（男）

（女）

两种情况的信息增量一样，因此要把两种情况都列出来。



Prepruning假设我们的深度是1，那么两种情况：



即使用单一特征做判断。

Postpruning基于验证集做判断：

1. 左树

未剪枝结果：正确、错误、正确、错误

将性别剪去结果：正确、错误、正确、正确

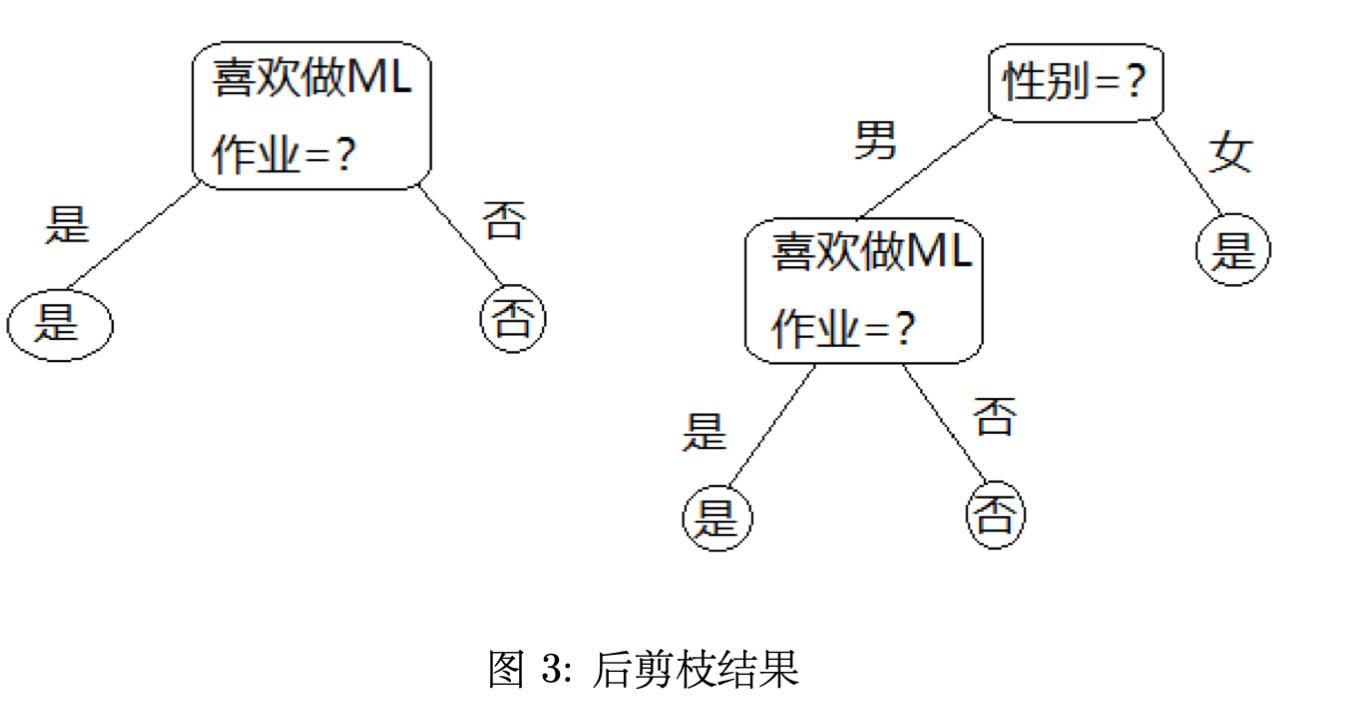
整体表现提升了，因此剪枝合理。

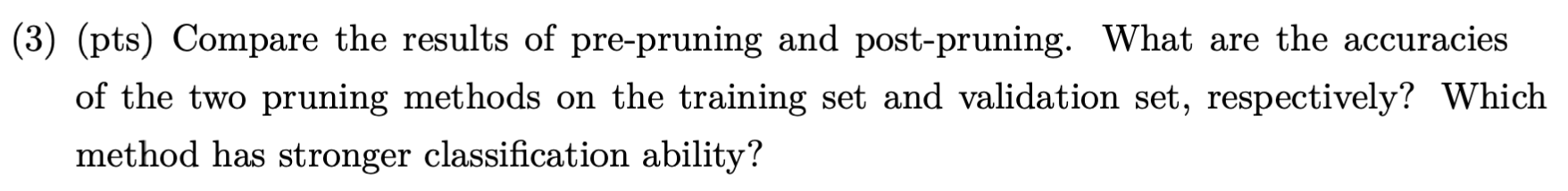
1. 右树

未剪枝结果：正确、错误、正确、错误

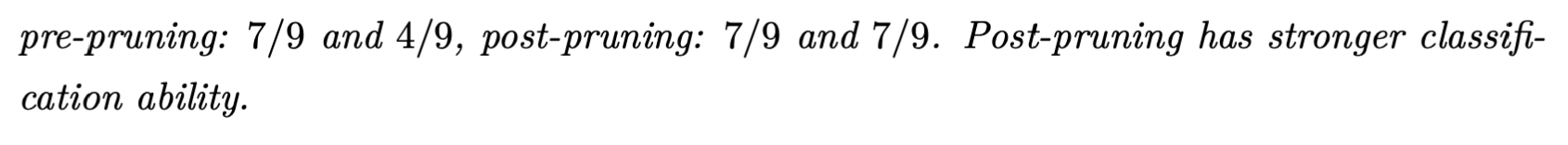
将ML成绩剪去结果：错误、错误、正确、错误

整体表现下降了，因此剪枝不合理。









注意的是要把两个数据集（训练和验证）都考虑进去。