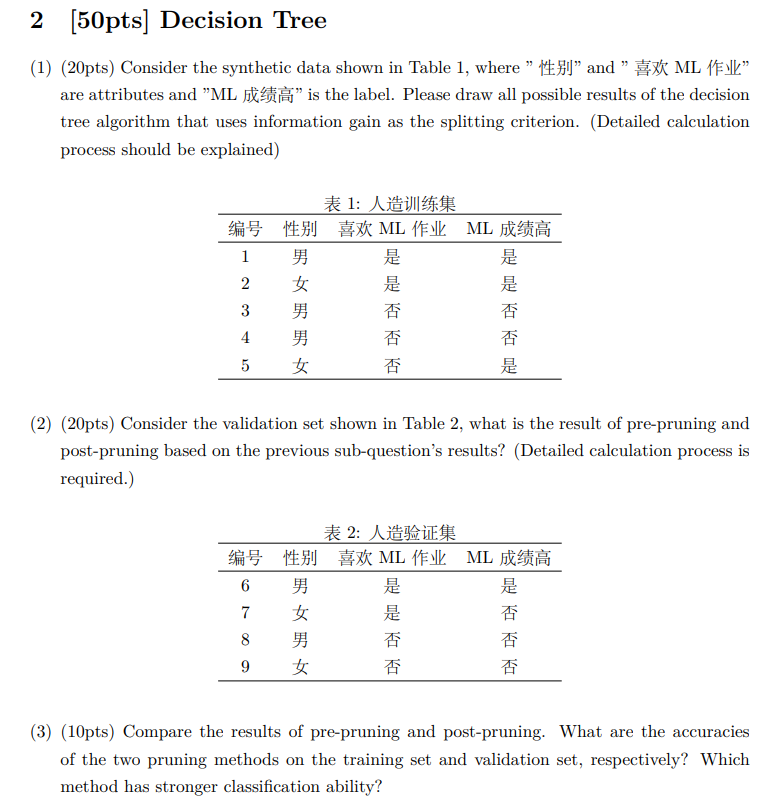


1. Answer:
2. All samples from current node belong to the same category.
3. Current attribute set is empty or all samples in the data set of the node take the same value
4. The data set of a child node is empty, then the child node is a leaf node.
5. We hope that the samples contained in the branch nodes of the decision tree belong to the same category as much as possible, that is, the purity of the nodes is getting higher and higher.

Suppose that the discrete feature a has possible values Then, splitting the data set by feature will produce child nodes, where the -th child node includes all samples in taking the value for feature . Then Information gain is:

Where , is the proportion of the -th class in the data set .

1. It might cause overfitting



1. Let’s calculate the information entropy of root node first. At the beginning of decision tree learning, the root node includes all examples in , where positive examples account for and negative examples account for . Thus, the information entropy is

Let and denote , , and , respectively. As we did when calculating , we can get

Therefore, the information gain of “性别” and “喜欢ML作业” is

Finally, we get the two possible decision trees

|  |  |
| --- | --- |
| Fig. 1 | Fig. 2 |

1. **Pre-pruning:**

Before splitting, all samples congregate at the root node. If we don’t split, the root node will be marked as a leaf node. Suppose we mark this node as “是”, then only one sample is classified correctly, so the accuracy is .

After splitting by attribute “喜欢ML作业”(shown in Fig. 3), there are 3 samples can be classified correctly, so the accuracy is . Therefore, we need such a splitting scheme.

Let’s consider on the next attribute “性别”. Before and after pruning, the accuracy is and respectively. Therefore, we don’t need to split.

We can also get the results corresponding to the decision tree in Fig. 2, as shown in Fig. 4

|  |  |
| --- | --- |
| Fig. 3 | Fig. 4 |

**Post-pruning:**

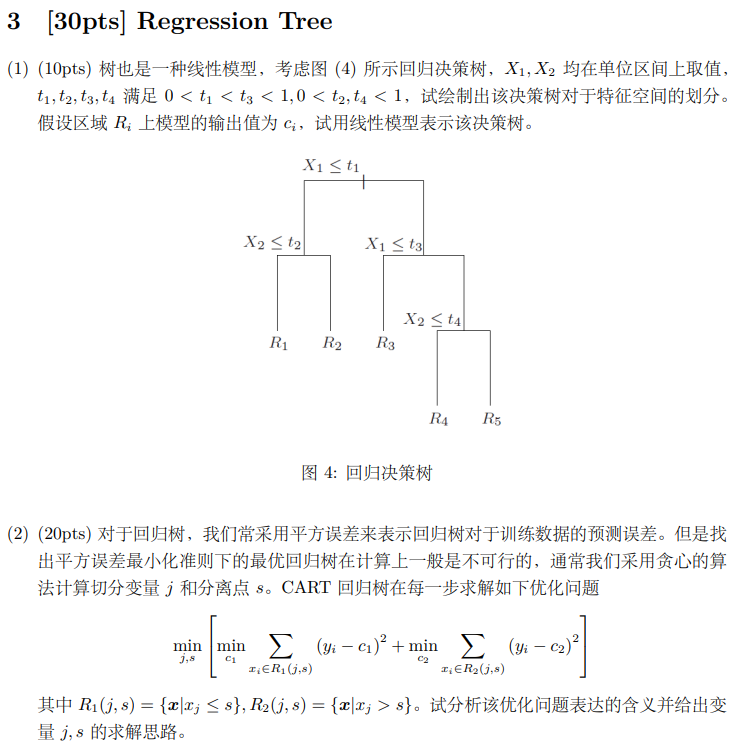
Let’s start from the decision tree shown in Fig. 1. Inspect the tree from the bottom up, if we keep “性别”, the accuracy will be , otherwise, the accuracy will be , so we have reason to cut this attribute. How about attribute “喜欢ML作业“ ? With it and without it, the accuracy is and respectively, thus, keep it.

We can also get the results corresponding to the decision tree in Fig. 2, as shown in Fig. 6

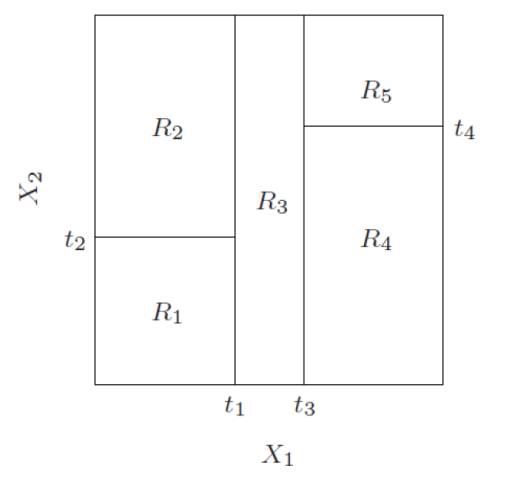
|  |  |
| --- | --- |
| Fig. 5 | Fig. 6 |

1. Post-pruning is better.

|  |  |  |
| --- | --- | --- |
| decision tree  Accuracy | Pre-pruning | Post-pruning |
| Fig. 1 | 7/9 | 7/9 |
| Fig. 2 | 4/9 | 7/9 |



1. According to the description, we can draw the feature space.



Decision tree:

1. This optimization problem solves the parameters so that the sum of the squared errors of the decision tree on the two sub-regions after splitting is the smallest.

Assuming that sets of data are input, each set of data has -dimensional features. When the value of the -th dimension feature is , traverse . When , traverses . Choose the parameter pair that minimizes . For the function , the minimum points of and are respectively , /