作业7

朱志儒 SA20225085

8.7（a）

（1）将每个元组的count作为计算属性选择方法中的一部分

（2）在决定元组多数表决的时候考虑元组的count

（b）

**选择根节点：**

Department:

|  |  |  |  |
| --- | --- | --- | --- |
| Sales | Senior | 30 | 110 |
| Junior | 80 |
| Systems | Senior | 8 | 31 |
| Junior | 23 |
| Marketing | Senior | 10 | 14 |
| Junior | 4 |
| Secretary | Senior | 4 | 10 |
| Junior | 6 |

Age

|  |  |  |  |
| --- | --- | --- | --- |
| 21…25 | Senior | 0 | 20 |
| Junior | 20 |
| 26…30 | Senior | 0 | 49 |
| Junior | 49 |
| 31…35 | Senior | 35 | 79 |
| Junior | 44 |
| 36…40 | Senior | 10 | 10 |
| Junior | 0 |
| 41…45 | Senior | 3 | 3 |
| Junior | 0 |
| 46…50 | Senior | 4 | 4 |
| Junior | 0 |

Salary

|  |  |  |  |
| --- | --- | --- | --- |
| 26K…30K | Senior | 0 | 46 |
| Junior | 46 |
| 31K…35K | Senior | 0 | 40 |
| Junior | 40 |
| 36K…40K | Senior | 4 | 4 |
| Junior | 0 |
| 41K…45K | Senior | 0 | 4 |
| Junior | 4 |
| 46K…50K | Senior | 40 | 63 |
| Junior | 23 |
| 66K…70K | Senior | 8 | 8 |
| Junior | 0 |

显然Salary的信息增益最大，选择Salary作为根节点。

**第二次划分：**

Department:

|  |  |  |  |
| --- | --- | --- | --- |
| Sales | Senior | 30 | 30 |
| Junior | 0 |
| Systems | Senior | 0 | 23 |
| Junior | 23 |
| Marketing | Senior | 10 | 10 |
| Junior | 0 |
| Secretary | Senior | 0 | 0 |
| Junior | 0 |

Age:

|  |  |  |  |
| --- | --- | --- | --- |
| 21…25 | Senior | 0 | 20 |
| Junior | 20 |
| 26…30 | Senior | 0 | 3 |
| Junior | 3 |
| 31…35 | Senior | 30 | 30 |
| Junior | 0 |
| 36…40 | Senior | 10 | 10 |
| Junior | 0 |
| 41…45 | Senior | 0 | 0 |
| Junior | 0 |
| 46…50 | Senior | 0 | 0 |
| Junior | 0 |

显然，department和age的信息增益相同，选择任意一个进行划分。

最后的决策树如下：

