**Задание**

Построить дерево принятия решений с использованием критерия Джини

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 0,1 | 10,1 | 1,4 | 0 |
| 7 | 45 | 1,7 | 0 |
| 77 | 9,2 | 19 | 1 |
| 1,2 | -70,4 | 48,5 | 0 |
| 8 | 45,2 | 18 | 1 |
| 12 | 19,8 | 14 | 0 |
| 45 | 14,1 | 99,1 | 0 |
| 200 | -700 | 99,7 | 0 |
| 48 | 99 | -8,1 | 1 |

Определить результат классификации для следующих данных:

|  |  |  |
| --- | --- | --- |
| **X1** | **X2** | **X3** |
| -70 | -80 | 1,12 |
| -4 | 100 | -7,4 |
| 80 | -44,7 | 1 |
| 64,2 | 102,4 | 14 |

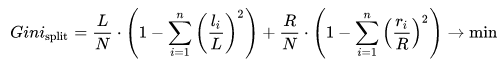
**Решение**

Поскольку вы сказали, что нужно использовать жадный алгоритм с критерием Джинни, я вышел на алгоритм CART, который заключает в себе жадный алгоритм с критерием Джинни  
По этому алгоритму необходимо вычислить критерий Джини для каждого разбиения в таблице и выбрать минимальное найденное значение для наиболее эффективного ветвления

 для разбиения на две части.



Тогда:



Эту формулу можно преобразовать к виду:

N — число примеров в узле — предке, L, R — число примеров соответственно в левом и правом потомке, li и ri — число экземпляров i-го класса в левом/правом потомке, n – количество классов.

В нашей задаче N = 9, n = 2.

Критерии разбиений:

X1>=65 Gini = 5 - ((1^2 + 2^2)/3 + (1^2 + 1^2)/2) = 6.3

X1>=3.4 Gini = 5 - ((0^2 + 0^2)/0 + (2^2 + 3^2)/5) = деление на ноль

X1>=3.4 Gini = 5 - ((0^2 + 0^2)/0 + (2^2 + 3^2)/5) = деление на ноль

X1>=1000 Gini = 5 - ((1^2 + 3^2)/4 + (1^2 + 0^2)/1) = 5.5

X1>=45 Gini = 5 - ((1^2 + 1^2)/2 + (1^2 + 2^2)/3) = 6.3

**X2>=4.1 Gini = 5 - ((2^2 + 0^2)/2 + (0^2 + 3^2)/3) = 4.0**

X2>=-40 Gini = 5 - ((1^2 + 0^2)/1 + (1^2 + 3^2)/4) = 5.5

X2>=80 Gini = 5 - ((2^2 + 2^2)/4 + (0^2 + 1^2)/1) = 6.0

X2>=-1000 Gini = 5 - ((0^2 + 0^2)/0 + (2^2 + 3^2)/5) = деление на ноль

X2>=46 Gini = 5 - ((2^2 + 1^2)/3 + (0^2 + 2^2)/2) = 5.3

X3>=14 Gini = 5 - ((2^2 + 2^2)/4 + (0^2 + 1^2)/1) = 6.0

X3>=8 Gini = 5 - ((1^2 + 2^2)/3 + (1^2 + 1^2)/2) = 6.3

X3>=1 Gini = 5 - ((0^2 + 1^2)/1 + (2^2 + 2^2)/4) = 6.0

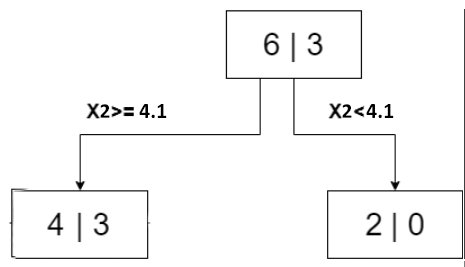
X3>=1 Gini = 5 - ((0^2 + 1^2)/1 + (2^2 + 2^2)/4) = 6.0

X3>=-7.1 Gini = 5 - ((0^2 + 0^2)/0 + (2^2 + 3^2)/5) = деление на ноль

Минимальное значения критерия Джини = 4.0

Это значение соответствует разбиению(**X1>=4.1**)

При ветвлении будем использовать правила слева записывать значения первого класса(Y=0), а справа для второго (Y=1)



**X2>=4.1**

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 7 | 45 | 1,7 | 0 |
| 77 | 9,2 | 19 | 1 |
| 8 | 45,2 | 18 | 1 |
| 12 | 19,8 | 14 | 0 |
| 45 | 14,1 | 99,1 | 0 |
| 200 | -700 | 99,7 | 0 |
| 48 | 99 | -8,1 | 1 |

**X2<4.1**

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 0,1 | 10,1 | 1,4 | 0 |
| 1,2 | -70,4 | 48,5 | 0 |

В таблице (**X1>=4.1**) можно найти локально оптимальное решение, его можно получить повторным нахождением критериев Джинни или можем увидеть, что при всех отрицательных X2 Y = 0 поэтому возьмем минимальное из возможных - 4.1  
  
Проверим нахождением критерия Джини  
X1>=7 Gini = 5 - ((0^2 + 0^2)/0 + (3^2 + 2^2)/5) = деление на ноль

X1>=77 Gini = 5 - ((2^2 + 2^2)/4 + (1^2 + 0^2)/1) = 6.0

X1>=8 Gini = 5 - ((1^2 + 0^2)/1 + (2^2 + 2^2)/4) = 6.0

X1>=12 Gini = 5 - ((1^2 + 1^2)/2 + (2^2 + 1^2)/3) = 6.3

X1>=45 Gini = 5 - ((2^2 + 1^2)/3 + (1^2 + 1^2)/2) = 6.3

X2>=45 Gini = 5 - ((2^2 + 1^2)/3 + (1^2 + 1^2)/2) = 6.3

X2>=9.2 Gini = 5 - ((0^2 + 0^2)/0 + (3^2 + 2^2)/5) = деление на ноль

X2>=45.2 Gini = 5 - ((3^2 + 1^2)/4 + (0^2 + 1^2)/1) = 5.5

X2>=19.8 Gini = 5 - ((1^2 + 1^2)/2 + (2^2 + 1^2)/3) = 6.3

X2>=14.1 Gini = 5 - ((1^2 + 0^2)/1 + (2^2 + 2^2)/4) = 6.0

X3>=1.7 Gini = 5 - ((0^2 + 0^2)/0 + (3^2 + 2^2)/5) = деление на ноль

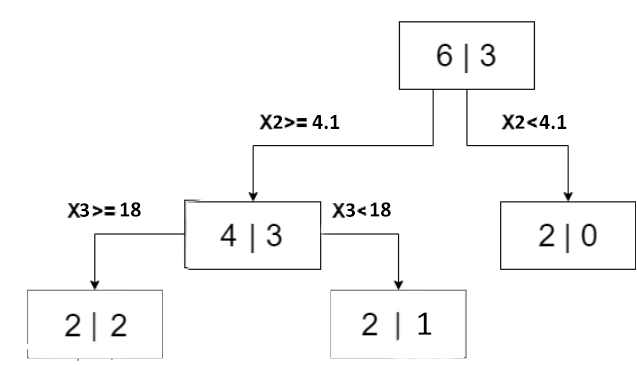
X3>=19 Gini = 5 - ((2^2 + 1^2)/3 + (1^2 + 1^2)/2) = 6.3

**X3>=18 Gini = 5 - ((2^2 + 0^2)/2 + (1^2 + 2^2)/3) = 5.3**

X3>=14 Gini = 5 - ((1^2 + 0^2)/1 + (2^2 + 2^2)/4) = 6.0

X3>=99.1 Gini = 5 - ((3^2 + 1^2)/4 + (0^2 + 1^2)/1) = 5.5

Продолжим построение дерева



**X3>=18**

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 77 | 9,2 | 19 | 1 |
| 8 | 45,2 | 18 | 1 |
| 45 | 14,1 | 99,1 | 0 |
| 200 | -700 | 99,7 | 0 |

X1>=7 Gini = 4 - ((0^2 + 0^2)/0 + (2^2 + 2^2)/4) = деление на ноль

X1>=77 Gini = 4 - ((1^2 + 1^2)/2 + (1^2 + 1^2)/2) = 7.0

X1>=45 Gini = 4 - ((0^2 + 1^2)/1 + (2^2 + 1^2)/3) = 6.3

X1>=200 Gini = 4 - ((1^2 + 2^2)/3 + (1^2 + 0^2)/1) = 6.3

X2>=9.2 Gini = 4 - ((1^2 + 0^2)/1 + (1^2 + 2^2)/3) = 6.3

X2>=45.2 Gini = 4 - ((2^2 + 1^2)/3 + (0^2 + 1^2)/1) = 6.3

X2>=14.1 Gini = 4 - ((1^2 + 1^2)/2 + (1^2 + 1^2)/2) = 7.0

X2>=-700 Gini = 4 - ((0^2 + 0^2)/0 + (2^2 + 2^2)/4) = деление на ноль

X3>=19 Gini = 4 - ((0^2 + 1^2)/1 + (2^2 + 1^2)/3) = 6.3

X3>=18 Gini = 4 - ((0^2 + 0^2)/0 + (2^2 + 2^2)/4) = деление на ноль

**X3>=99.1 Gini = 4 - ((0^2 + 2^2)/2 + (2^2 + 0^2)/2) = 5.0**

X3>=99.7 Gini = 4 - ((1^2 + 2^2)/3 + (1^2 + 0^2)/1) = 6.3

**X3<18**

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 7 | 45 | 1,7 | 0 |
| 12 | 19,8 | 14 | 0 |
| 48 | 99 | -8,1 | 1 |

X1>=7 Gini = 3 - ((0^2 + 0^2)/0 + (2^2 + 1^2)/3) = деление на ноль

X1>=12 Gini = 3 - ((1^2 + 0^2)/1 + (1^2 + 1^2)/2) = 7.0

**X1>=48 Gini = 3 - ((2^2 + 0^2)/2 + (0^2 + 1^2)/1) = 6.0**

X2>=45 Gini = 3 - ((1^2 + 0^2)/1 + (1^2 + 1^2)/2) = 7.0

X2>=19.8 Gini = 3 - ((0^2 + 0^2)/0 + (2^2 + 1^2)/3) = деление на ноль

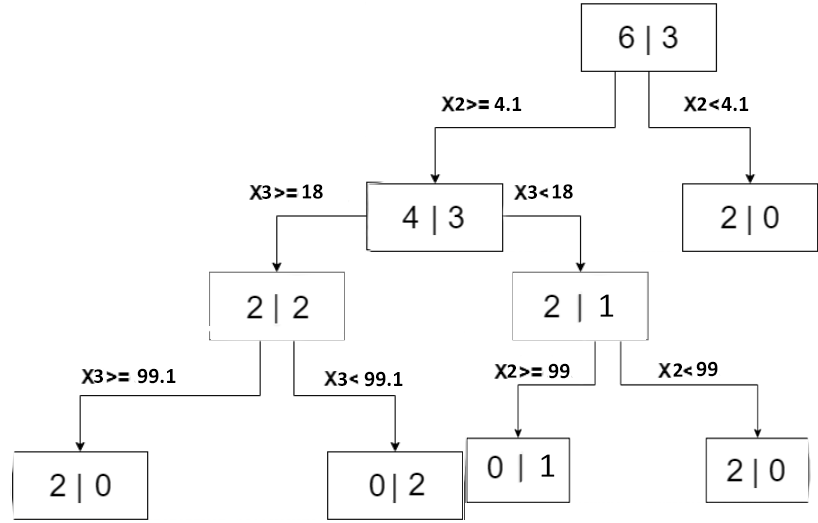
**X2>=99 Gini = 3 - ((2^2 + 0^2)/2 + (0^2 + 1^2)/1) = 6.0**

**X3>=1.7 Gini = 3 - ((0^2 + 1^2)/1 + (2^2 + 0^2)/2) = 6.0**

X3>=14 Gini = 3 - ((1^2 + 1^2)/2 + (1^2 + 0^2)/1) = 7.0

X3>=-8.1 Gini = 3 - ((0^2 + 0^2)/0 + (2^2 + 1^2)/3) = деление на ноль

Продолжим построение дерева



Классификация выполняется в соответствии с выделенными правилами:

Y = 0, если

(x2 ≥ 4.1 && x3 ≥ 18 && x3≥ 99.1) ||

(x2≥ 4.1 && x3 < 18 && x2 < 99) ||

(x2 < 4.1),иначе Y = 1

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| -70 | -80 | 1,12 | 0 |
| -4 | 100 | -7,4 | 1 |
| 80 | -44,7 | 1 | 0 |
| 64,2 | 102,4 | 14 | 1 |