**Задание**

Построить дерево принятия решений с использованием критерия Джини

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| -80,1 | -4 | 12,5 | 0 |
| 0 | 99 | 78,4 | 0 |
| 65 | 4,1 | 14 | 1 |
| 3,4 | -40 | 8 | 0 |
| 3,4 | 80 | 1 | 1 |
| 2,2 | 18,1 | 12 | 0 |
| 0 | 15 | 47,1 | 0 |
| 1000 | -1000 | 1 | 0 |
| 45 | 46 | -7,1 | 1 |

Определить результат классификации для следующих данных:

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

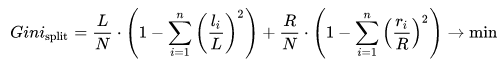
**Решение**

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>=-80.1 Gini = 9 - ((0^2 + 0^2)/0 + (6^2 + 3^2)/9) = div by zero

X1>=0 Gini = 9 - ((1^2 + 0^2)/1 + (5^2 + 3^2)/8) = 3.8

X1>=65 Gini = 9 - ((5^2 + 2^2)/7 + (1^2 + 1^2)/2) = 3.9

**X1>=3.4 Gini = 9 - ((4^2 + 0^2)/4 + (2^2 + 3^2)/5) = 2.4**

**X1>=3.4 Gini = 9 - ((4^2 + 0^2)/4 + (2^2 + 3^2)/5) = 2.4**

X1>=2.2 Gini = 9 - ((3^2 + 0^2)/3 + (3^2 + 3^2)/6) = 3.0

X1>=0 Gini = 9 - ((1^2 + 0^2)/1 + (5^2 + 3^2)/8) = 3.8

X1>=1000 Gini = 9 - ((5^2 + 3^2)/8 + (1^2 + 0^2)/1) = 3.8

X1>=45 Gini = 9 - ((5^2 + 1^2)/6 + (1^2 + 2^2)/3) = 3.0

X2>=-4 Gini = 9 - ((2^2 + 0^2)/2 + (4^2 + 3^2)/7) = 3.4

X2>=99 Gini = 9 - ((5^2 + 3^2)/8 + (1^2 + 0^2)/1) = 3.8

X2>=4.1 Gini = 9 - ((3^2 + 0^2)/3 + (3^2 + 3^2)/6) = 3.0

X2>=-40 Gini = 9 - ((1^2 + 0^2)/1 + (5^2 + 3^2)/8) = 3.8

X2>=80 Gini = 9 - ((5^2 + 2^2)/7 + (1^2 + 1^2)/2) = 3.9

X2>=18.1 Gini = 9 - ((4^2 + 1^2)/5 + (2^2 + 2^2)/4) = 3.6

X2>=15 Gini = 9 - ((3^2 + 1^2)/4 + (3^2 + 2^2)/5) = 3.9

X2>=-1000 Gini = 9 - ((0^2 + 0^2)/0 + (6^2 + 3^2)/9) = div by zero

X2>=46 Gini = 9 - ((5^2 + 1^2)/6 + (1^2 + 2^2)/3) = 3.0

X3>=12.5 Gini = 9 - ((3^2 + 2^2)/5 + (3^2 + 1^2)/4) = 3.9

X3>=78.4 Gini = 9 - ((5^2 + 3^2)/8 + (1^2 + 0^2)/1) = 3.8

X3>=14 Gini = 9 - ((4^2 + 2^2)/6 + (2^2 + 1^2)/3) = div by zero

X3>=8 Gini = 9 - ((1^2 + 2^2)/3 + (5^2 + 1^2)/6) = 3.0

X3>=1 Gini = 9 - ((0^2 + 1^2)/1 + (6^2 + 2^2)/8) = 3.0

X3>=12 Gini = 9 - ((2^2 + 2^2)/4 + (4^2 + 1^2)/5) = 3.6

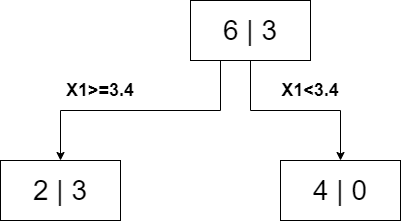
X3>=47.1 Gini = 9 - ((4^2 + 3^2)/7 + (2^2 + 0^2)/2) = 3.4

X3>=1 Gini = 9 - ((0^2 + 1^2)/1 + (6^2 + 2^2)/8) = 3.0

X3>=-7.1 Gini = 9 - ((0^2 + 0^2)/0 + (6^2 + 3^2)/9) = div by zero

Минимальное значения критерия Джини = 2.40   
Это значение соответствует двум равным разбиениям(**X1>=3.4**), поэтому выберем любое

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



**X1>=3.4**

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 65 | 4,1 | 14 | 1 |
| 3,4 | -40 | 8 | 0 |
| 3,4 | 80 | 1 | 1 |
| 1000 | -1000 | 1 | 0 |
| 45 | 46 | -7,1 | 1 |

**X1<3.4**

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| -80,1 | -4 | 12,5 | 0 |
| 0 | 99 | 78,4 | 0 |
| 2,2 | 18,1 | 12 | 0 |
| 0 | 15 | 47,1 | 0 |

В таблице (**X1>=3.4**) 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>=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) = div by zero

X1>=3.4 Gini = 5 - ((0^2 + 0^2)/0 + (2^2 + 3^2)/5) = div by zero

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) = div by zero

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

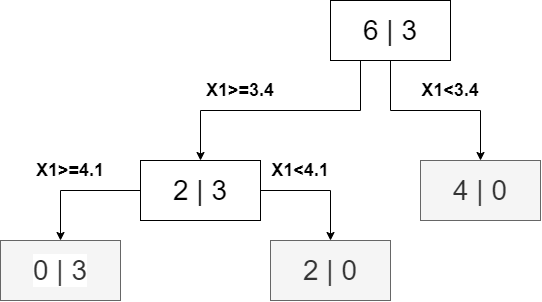
X3>=14 Gini = 5 - ((2^2 + 2^2)/4 + (0^2 + 1^2)/1) = 6.0

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) = div by zero

Продолжим построение дерева

Классификация выполняется в соответствии с выделенными правилами:

Y = 1, если x1 ≥ 3.4 и x2 ≥ 4.1, иначе Y = 0

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

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