**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 |

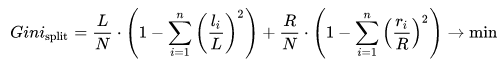
**Решение**

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>=0.1 Gini = 9 - ((0^2 + 0^2)/0 + (6^2 + 3^2)/9) = деление на ноль

X1>=7 Gini = 9 - ((2^2 + 0^2)/2 + (4^2 + 3^2)/7) = 3.4

X1>=77 Gini = 9 - ((5^2 + 2^2)/7 + (1^2 + 1^2)/2) = 3.9

X1>=1.2 Gini = 9 - ((1^2 + 0^2)/1 + (5^2 + 3^2)/8) = 3.8

X1>=8 Gini = 9 - ((3^2 + 0^2)/3 + (3^2 + 3^2)/6) = 3.0

X1>=12 Gini = 9 - ((3^2 + 1^2)/4 + (3^2 + 2^2)/5) = 3.9

X1>=45 Gini = 9 - ((4^2 + 1^2)/5 + (2^2 + 2^2)/4) = 3.6

X1>=200 Gini = 9 - ((5^2 + 3^2)/8 + (1^2 + 0^2)/1) = 3.8

X1>=48 Gini = 9 - ((5^2 + 1^2)/6 + (1^2 + 2^2)/3) = 3.0

X2>=10.1 Gini = 9 - ((2^2 + 1^2)/3 + (4^2 + 2^2)/6) = 4.0

X2>=45 Gini = 9 - ((5^2 + 1^2)/6 + (1^2 + 2^2)/3) = 3.0

X2>=9.2 Gini = 9 - ((2^2 + 0^2)/2 + (4^2 + 3^2)/7) = 3.4

X2>=-70.4 Gini = 9 - ((1^2 + 0^2)/1 + (5^2 + 3^2)/8) = 3.8

**X2>=45.2 Gini = 9 - ((6^2 + 1^2)/7 + (0^2 + 2^2)/2) = 1.7**

X2>=19.8 Gini = 9 - ((4^2 + 1^2)/5 + (2^2 + 2^2)/4) = 3.6

X2>=14.1 Gini = 9 - ((3^2 + 1^2)/4 + (3^2 + 2^2)/5) = 3.9

X2>=-700 Gini = 9 - ((0^2 + 0^2)/0 + (6^2 + 3^2)/9) = деление на ноль

X2>=99 Gini = 9 - ((6^2 + 2^2)/8 + (0^2 + 1^2)/1) = 3.0

X3>=1.4 Gini = 9 - ((0^2 + 1^2)/1 + (6^2 + 2^2)/8) = 3.0

X3>=1.7 Gini = 9 - ((1^2 + 1^2)/2 + (5^2 + 2^2)/7) = 3.9

X3>=19 Gini = 9 - ((3^2 + 2^2)/5 + (3^2 + 1^2)/4) = 3.9

X3>=48.5 Gini = 9 - ((3^2 + 3^2)/6 + (3^2 + 0^2)/3) = 3.0

X3>=18 Gini = 9 - ((3^2 + 1^2)/4 + (3^2 + 2^2)/5) = 3.9

X3>=14 Gini = 9 - ((2^2 + 1^2)/3 + (4^2 + 2^2)/6) = 4.0

X3>=99.1 Gini = 9 - ((4^2 + 3^2)/7 + (2^2 + 0^2)/2) = 3.4

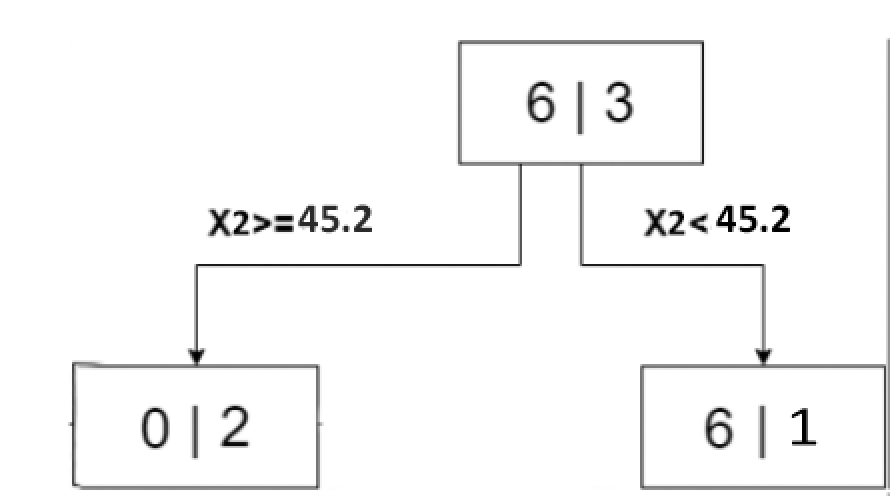
X3>=99.7 Gini = 9 - ((5^2 + 3^2)/8 + (1^2 + 0^2)/1) = 3.8

X3>=-8.1 Gini = 9 - ((0^2 + 0^2)/0 + (6^2 + 3^2)/9) = деление на ноль

Минимальное значения критерия Джини = 1.7

Это значение соответствует разбиению(**X2>=45.2**)

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



**X2>=42.5**

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 8 | 45,2 | 18 | 1 |
| 48 | 99 | -8,1 | 1 |

**X2<42.5**

|  |  |  |  |
| --- | --- | --- | --- |
| **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 |
| 12 | 19,8 | 14 | 0 |
| 45 | 14,1 | 99,1 | 0 |
| 200 | -700 | 99,7 | 0 |

В таблице (**X2<42.5**) можно найти локально оптимальное решение, его можно получить повторным нахождением критериев Джинни.

Проверим нахождением критерия Джини  
X1>=0.1 Gini = 7 - ((0^2 + 0^2)/0 + (6^2 + 1^2)/7) = деление на ноль

X1>=7 Gini = 7 - ((2^2 + 0^2)/2 + (4^2 + 1^2)/5) = 3.6

**X1>=77 Gini = 7 - ((5^2 + 0^2)/5 + (1^2 + 1^2)/2) = 3.0**

X1>=1.2 Gini = 7 - ((1^2 + 0^2)/1 + (5^2 + 1^2)/6) = 3.7

X1>=12 Gini = 7 - ((3^2 + 0^2)/3 + (3^2 + 1^2)/4) = 3.5

X1>=45 Gini = 7 - ((4^2 + 0^2)/4 + (2^2 + 1^2)/3) = 3.3

X1>=200 Gini = 7 - ((5^2 + 1^2)/6 + (1^2 + 0^2)/1) = 3.7

X2>=10.1 Gini = 7 - ((2^2 + 1^2)/3 + (4^2 + 0^2)/4) = 3.3

X2>=45 Gini = 7 - ((5^2 + 1^2)/6 + (1^2 + 0^2)/1) = 3.7

X2>=9.2 Gini = 7 - ((2^2 + 0^2)/2 + (4^2 + 1^2)/5) = 3.6

X2>=-70.4 Gini = 7 - ((1^2 + 0^2)/1 + (5^2 + 1^2)/6) = 3.7

X2>=19.8 Gini = 7 - ((4^2 + 1^2)/5 + (2^2 + 0^2)/2) = 3.6

X2>=14.1 Gini = 7 - ((3^2 + 1^2)/4 + (3^2 + 0^2)/3) = 3.5

X2>=-700 Gini = 7 - ((0^2 + 0^2)/0 + (6^2 + 1^2)/7) = деление на ноль

X3>=1.4 Gini = 7 - ((0^2 + 0^2)/0 + (6^2 + 1^2)/7) = деление на ноль

X3>=1.7 Gini = 7 - ((1^2 + 0^2)/1 + (5^2 + 1^2)/6) = 3.7

X3>=19 Gini = 7 - ((3^2 + 0^2)/3 + (3^2 + 1^2)/4) = 3.5

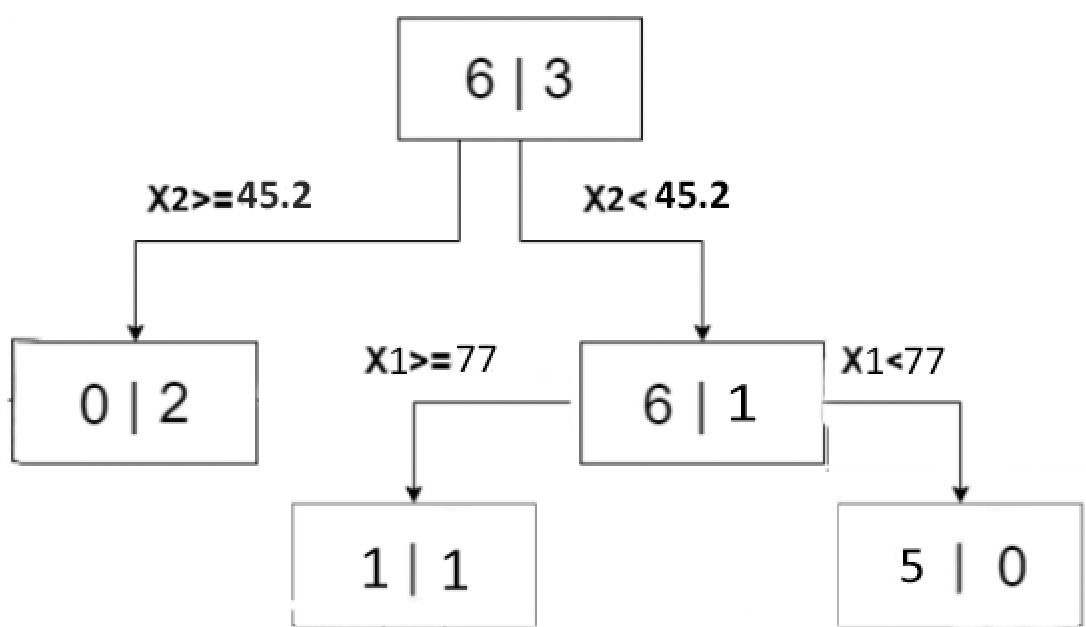
X3>=48.5 Gini = 7 - ((3^2 + 1^2)/4 + (3^2 + 0^2)/3) = 3.5

X3>=14 Gini = 7 - ((2^2 + 0^2)/2 + (4^2 + 1^2)/5) = 3.6

X3>=99.1 Gini = 7 - ((4^2 + 1^2)/5 + (2^2 + 0^2)/2) = 3.6

X3>=99.7 Gini = 7 - ((5^2 + 1^2)/6 + (1^2 + 0^2)/1) = 3.7

Продолжим построение дерева



**X1>=77**

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

**X1<77**

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

В таблице (**X1<77**) можно найти локально оптимальное решение, его можно получить повторным нахождением критериев Джинни.

X1>=77 Gini = 2 - ((0^2 + 0^2)/0 + (1^2 + 1^2)/2) = деление на ноль

X1>=200 Gini = 2 - ((0^2 + 1^2)/1 + (1^2 + 0^2)/1) = 7.0

**X2>=9.2 Gini = 2 - ((1^2 + 0^2)/1 + (0^2 + 1^2)/1) = 7.0**

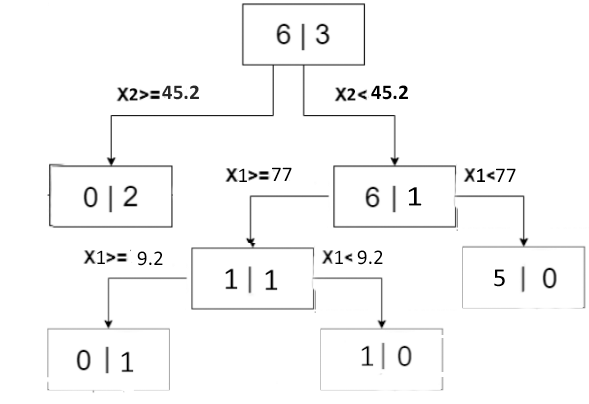
X2>=-700 Gini = 2 - ((0^2 + 0^2)/0 + (1^2 + 1^2)/2) = деление на ноль

X3>=19 Gini = 2 - ((0^2 + 0^2)/0 + (1^2 + 1^2)/2) = деление на ноль

X3>=99.7 Gini = 2 - ((0^2 + 1^2)/1 + (1^2 + 0^2)/1) = 7.0

У меня три одинаковых результата,выберу один на выбор.

Продолжим построение



Классификация выполняется в соответствии с выделенными правилами:

Y = 0, если

(x2 < 42.5 && x1 ≥ 77 && x1< 9.2) ||

(x2< 42.5 && x1 < 77) ,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 |