**Задание**

Построить дерево принятия решений с использованием критерия Джини

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 37,1 | 3,7 | 98,3 | 0 |
| 26,4 | 3,8 | -80,2 | 0 |
| 55,3 | -4,5 | 87,4 | 1 |
| -40,5 | -100,8 | 11,3 | 0 |
| -78,3 | 0,09 | 11,2 | 1 |
| -90,4 | -7,9 | 30,6 | 0 |
| 40,2 | 0 | -56,3 | 0 |
| 50,2 | -11,4 | 80,2 | 0 |
| 60,7 | -4,3 | 17,5 | 1 |

Определить результат классификации для следующих данных:

|  |  |  |
| --- | --- | --- |
| **X1** | **X2** | **X3** |
| 20,4 | -11 | 80,5 |
| 30,1 | -8 | 44,7 |
| 44,2 | -15,2 | -20 |
| 70,8 | -7 | -98,2 |

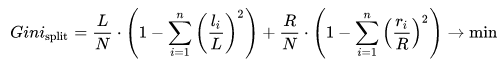
**Решение**

Следуя жадному алгоритму, мы должны выбирать локально оптимальные решения. То есть необходимо вычислить критерий Джини каждого разбиения. Он определяется:

 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 ≥ -90.4 Gini = 9 – ((0 + 0)/0 +… =/0

X1 ≥ -78.3 (попало 9 чисел) Gini = 9 – ((0 + 1)/1(кол-во сверху) + (5^2 + 3^2)/8) =3.75, где 5=нолики, 3= единички)

X1 ≥ -40.5 Gini = 9 – ((1^2 + 1^2)/2 + (5^2 + 2^2)/7) = 3.86

X1 ≥ 26.4 Gini = 9 – ((2^2 + 1^2)/3 + (4^2 + 2^2)/6) = 4

X1 ≥ 37.1 Gini = 9 – ((3^2 + 1^2)/4 + (3^2 + 2^2)/5) =3.9

X1 ≥ 40.2 Gini = 9 – ((4^2 + 1^2)/5 + (2^2 + 2^2)/4) = 3.6

X1 ≥ 50.2 Gini = 9 – ((5^2 + 1^2)/6 + (1^2 + 2^2)/3) = 3

**X1 ≥ 55.3 Gini = 9 – ((6^2 + 1^2)/7 + (0 + 2^2)/2) = 9 –(5.285+2)=1.7**

X1 ≥ 60.7 Gini = 9 – ((6^2 + 2^2)/8 + (1 + 0)/1) = 3

X2 ≥ -100.8 Gini = 9 – ((0 + 0)/0 + … = /0

X2 ≥ -11.4 Gini = 9 – ((1 + 0)/1 + (5^2 + 3^2)/8) =3.75

X2 ≥ -7.9 Gini = 9 – ((2^2 + 0)/2 + (4^2 + 3^2)/7) = 3.43

X2 ≥ -4.5 Gini = 9 – ((3^2 + 0)/3 + (3^2 + 3^2)/6) =3

X2 ≥ -4.3 Gini = 9 – ((3^2 + 1)/4 + (3^2 + 2^2)/5) =3,9

X2 ≥ 0 Gini = 9 – ((3^2 + 2^2)/5 + (3^2 + 1^2)/4) = 3.9

X2 ≥ 0.09 Gini = 9 – ((4^2 + 2^2)/6 + (2^2 + 1)/3) =4

X2 ≥ 3.7 Gini = 9 – ((4^2 + 3^2)/7 + (2^2 + 0)/2) = 3.43

X2 ≥ 3.8 Gini = 9 – ((5^2 + 3^2)/8 + (1 + 0)/1) = 9-(4.25+1) = 3.75

X3 ≥ -80.2 Gini = 9 – ((0 + 0)/0 +… = /0

X3 ≥ -56.3 Gini = 9 – ((1 + 0)/1 + (5^2 + 3^2)/8) = 3.75

X3 ≥ 11.2 Gini = 9 – ((2^2 + 0)/2 + (4^2 + 3^2)/7) = 3.43

X3 ≥ 11.3 Gini = 9 – ((2^2 + 1^2)/3 + (4^2 + 2^2)/6) = 4

X3 ≥ 17.5 Gini = 9 – ((3^2 + 1^2)/4 + (3^2 + 2^2)/5) = 3,9

X3 ≥ 30.6 Gini = 9 – ((3^2 + 2^2)/5 + (3^2 + 1^2)/4) = 3.9

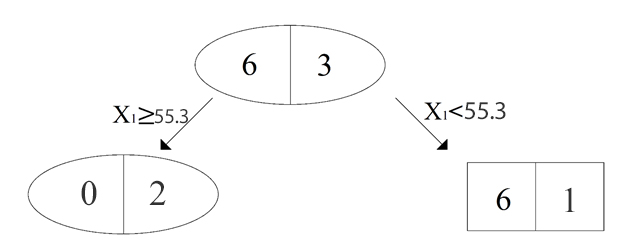
X3 ≥ 80.2 Gini = 9 – ((4^2 + 2^2)/6+ (2^2 + 1^2)/3) = 4

X3 ≥ 87.4 Gini = 9 – ((5^2 + 2^2)/7+ (1^2 + 1)/2) = 3,857

X3 ≥ 98.3 Gini = 9 – ((5^2 + 3^2)/8+ (1^2 + 0)/1) = 3.75

Минимальное значения критерия Джини = 1.7, соответствующее разбиению x1 ≥ 55.3

При ветвлении с одной стороны был получен набор только из элементов первого класса, с другой – шести первого и одного второго.



|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| -90,4 | -7,9 | 30,6 | 0 |
| -78,3 | 0,09 | 11,2 | 1 |
| -40,5 | -100,8 | 11,3 | 0 |
| 26,4 | 3,8 | -80,2 | 0 |
| 37,1 | 3,7 | 98,3 | 0 |
| 40,2 | 0 | -56,3 | 0 |
| 50,2 | -11,4 | 80,2 | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Y** |
| 55,3 | -4,5 | 87,4 | 1 |
| 60,7 | -4,3 | 17,5 | 1 |

В таблице (X1>=55.3) можно найти локально оптимальное решение, его можно получить пповторным нахождением критериев Джини или можем увидеть, что при всех отрицательных X2 Y = 0 поэтому возьмем минимальное из возможных - 4.1

Проверим нахождением критерия Джини

X1>=-90.4 Gini = /0

X1>=-78.3 Gini = 7 - ((1^2 + 0^2)/1 + (5^2 + 1)/6) = 1.6667

X1>=-40.5 Gini = 7 - ((1^2 + 1^2)/2 + (5^2 + 0^2)/5) = 1

X1>=26.4 Gini = 7 - ((2^2 + 1^2)/3 + (4^2 + 0^2)/4) = 1.3333

X1>=37.1 Gini = 7 - ((3^2 + 1^2)/4 + (3^2 + 0^2)/3) = 7-(2.5+3)=1.5

X1>=40.2 Gini = 7 - ((4^2 + 1^2)/5 + (2^2 + 0^2)/2) = 7-(3.4+2)=1.6

X1>=50.2 Gini = 7 - ((5^2 + 1^2)/6 + (1^2 + 0^2)/1) = 7-(4.333+1)= 1.66667

X2>=-100.8 Gini = /0

X2>=-11.4 Gini = 7 - ((1^2 + 0^2)/1 + (5^2 + 1)/6) = 1.6667

X2>=-7.9 Gini = 7 - ((2^2 + 0^2)/2 + (4^2 + 1^2)/5) = 1.6

X2>=0 Gini = 7 - ((3^2 + 0^2)/3 + (3^2 + 1^2)/4) = 1.5

X2>=0.09 Gini = 7 - ((4^2 + 0^2)/4 + (2^2 + 1^2)/3) = 7-(2.5+3)= 1.3333

X2>=3.7 Gini = 7 - ((4^2 + 1^2)/5 + (2^2 + 0^2)/2) = 7-(3.4+2)=1.6

X2>3.8= Gini = 7 - ((5^2 + 1^2)/6 + (1^2 + 0^2)/1) = 7-(4.333+1)= 1.66667

X3>=-80.2 Gini = /0

X3>=-56.3 Gini = 7 - ((1^2 + 0^2)/1 + (5^2 + 1)/6) = 1.6667

X3>=11.2 Gini = 7 - ((2^2 + 0^2)/2 + (4^2 + 1^2)/5) = 1.6

X3>=11.3 Gini = 7 - ((2^2 + 1^2)/3 + (4^2 + 0^2)/4) = 1.3333

X3>=30.6 Gini = 7 - ((3^2 + 1^2)/4 + (3^2 + 0^2)/3) = 7-(2.5+3)=1.5

X3>=80.2 Gini = 7 - ((4^2 + 1^2)/5 + (2^2 + 0^2)/2) = 7-(3.4+2)=1.6

X3>=98.3 Gini = 7 - ((5^2 + 1^2)/6 + (1^2 + 0^2)/1) = 7-(4.333+1)= 1.66667

6|3

X1>=55.3 X1<55.3

6|1

0|2

X1<-40.5

X1>=40.5

5|0

1|1

X1>=-90.4 Gini = 2 - ((0^2 + 0^2)/0 + (1^2 + 1^2)/2) = деление на ноль

X1>=-78.3 Gini = 2 - ((0^2 + 1^2)/1 + (1^2 + 0^2)/1) = 7.0

X2>=-7.9 Gini = 2 - ((0^2 + 0^2)/0 + (1^2 + 1^2)/2) = деление на ноль

**X2>=0.09 Gini = 2 - ((1^2 + 0^2)/1 + (0^2 + 1^2)/1) = 7.0**

X3>=11.2 Gini = 2 - ((0^2 + 0^2)/0 + (1^2 + 1^2)/2) = деление на ноль

X3>=30.6 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

6|3

X1>=55.3 X1<55.3

6|1

0|2

X1<-40.5

X1>=-40.5

5|0

1|1

X2<0.09 X2>=0.09

1|0

0|1

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | Y |
| 20,4 | -11 | 80,5 | 0 |
| 30,1 | -8 | 44,7 | 0 |
| 44,2 | -15,2 | -20 | 0 |
| 70,8 | -7 | -98,2 | 1 |