

Задачи на построение решающих деревьев

1. Постройте регрессионное дерево для прогнозирования y с помощью x на обучающей выборке:

x_i	0	1	2	3
y_i	5	6	4	100

Критерий деления узла на два – минимизация MSE. Дерево строится до трёх терминальных узлов.

2. Приведите пример, когда индекс Джини равен 0, 0.5, 0.99.
3. Дон-Жуан предпочитает брюнеток. Перед Новым Годом он посчитал, что в записной книжке у него 20 блондинок, 40 брюнеток, 2 рыжих и 8 шатенок. С Нового Года Дон-Жуан решил перенести все сведения в две записные книжки, в одну – брюнеток, во вторую – остальных.
Как изменились индекс Джини и энтропия в результате такого разбиения?

Домашнее задание.

1. Постройте регрессионное дерево для прогнозирования y с помощью x на обучающей выборке:

y_i	x_i
100	1
102	2
103	3
50	4
55	5
61	6
70	7

Критерий деления узла на два – минимизация MSE. Узлы делятся до тех пор, пока в узле остаётся больше двух наблюдений.

2. Постройте пример, когда индекс Джини больше 0.9999.
3. Машка пять дней гадала на ромашке, а затем выкладывала очередную фотку “Машка с ромашкой” в инстаграмчик. Результат гадания – переменная y_i , количество лайков у фотки – переменная x_i . Постройте классификационное дерево для прогнозирования y_i с помощью x_i на обучающей выборке:

y_i	x_i
плюнет	10
поцелует	11
поцелует	12
к сердцу прижмёт	13
к сердцу прижмёт	14

Дерево строится до идеальной классификации. Критерий деления узла на два – максимальное падение индекса Джини.