

## Задачи по решающим деревьям.

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  $RSS$ . Дерево строится до трёх терминальных узлов.

2.

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  $RSS$ . Узлы делятся до тех пор, пока в узле остаётся больше двух наблюдений.

3.

Дон-Жуан предпочитает брюнеток. Перед Новым Годом он посчитал, что в записной книжке у него 20 блондинок, 40 брюнеток, две рыжих и восемь 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?

4.

Приведите примеры наборов данных, для которых индекс Джини равен 0, 0.5 и 0.999.

5.

Машка пять дней подряд 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.