Задачи на построение решающих деревьев

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



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

1. Приведите пример, когда индекс Джини равен 0, 0.5, 0.99.
2. 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:



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

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



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