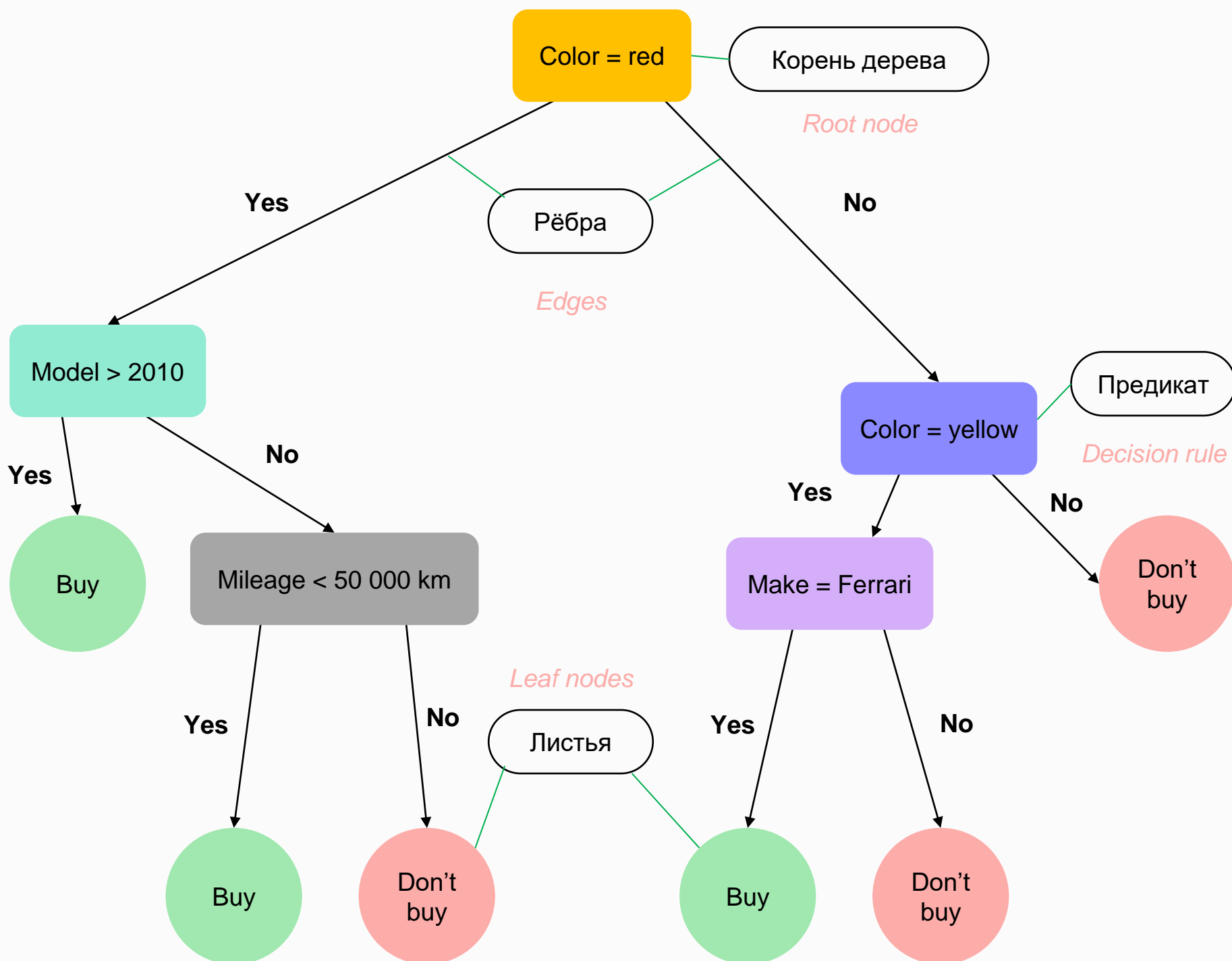


Решающее дерево. Критерии ветвления

Елена Кантонистова

Skillbox

Терминология



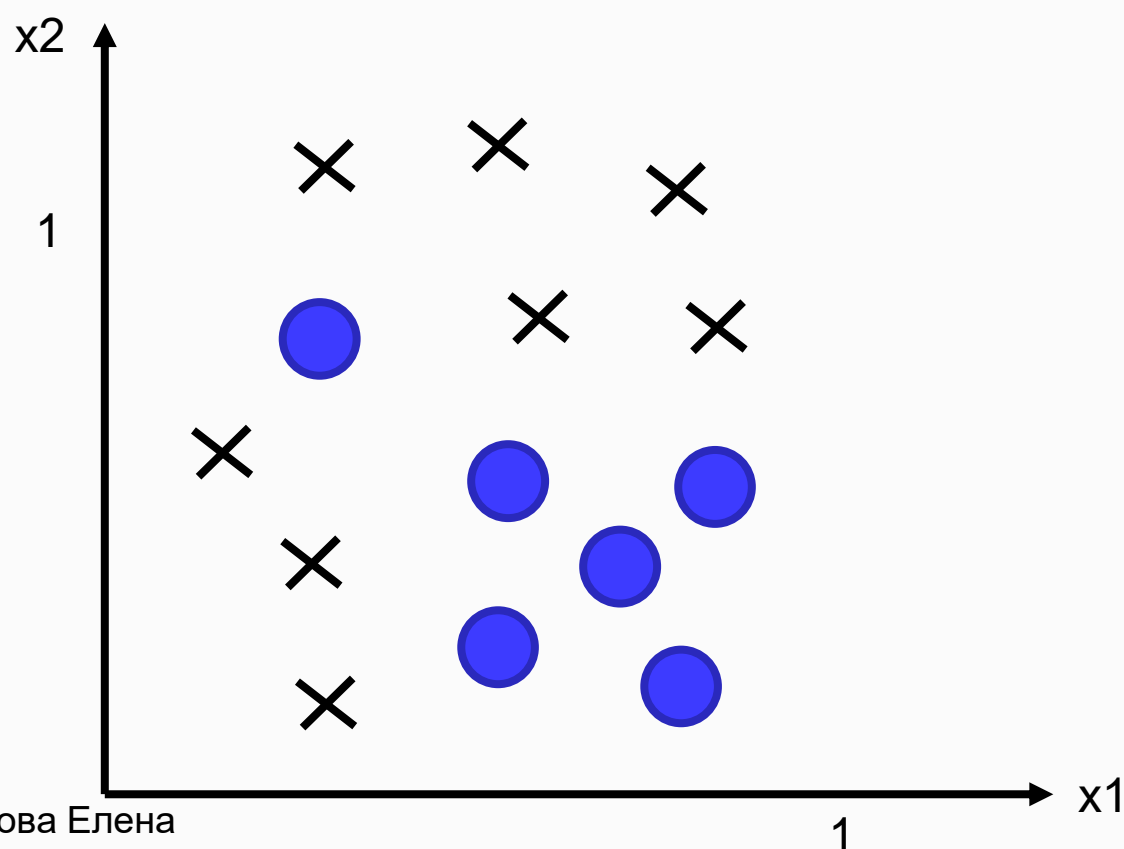
Предикаты

- Вы будете работать только с предикатами, в которых используется только один признак объекта, и он сравнивается с некоторым пороговым значением



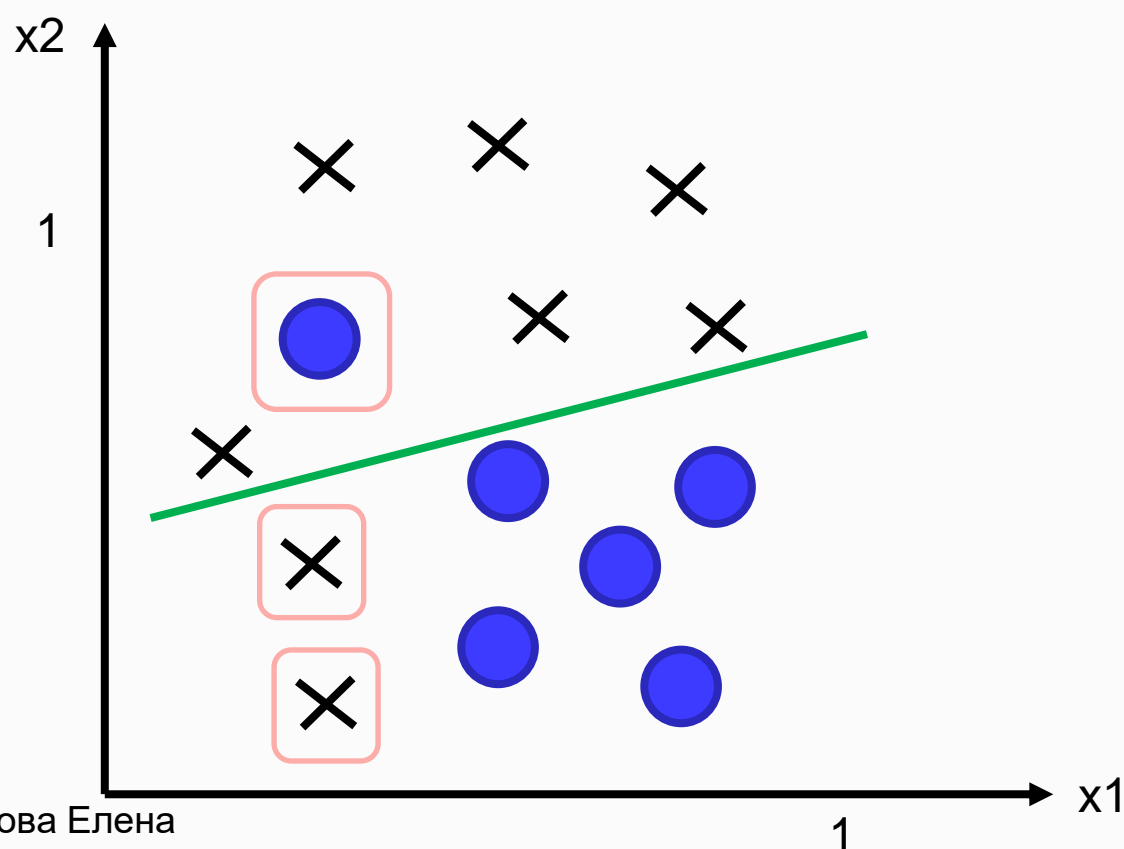
Пример: бинарная классификация

- Объекты двух классов — крестики и нолики
- У каждого объекта два признака — его координаты на плоскости x_1 и x_2



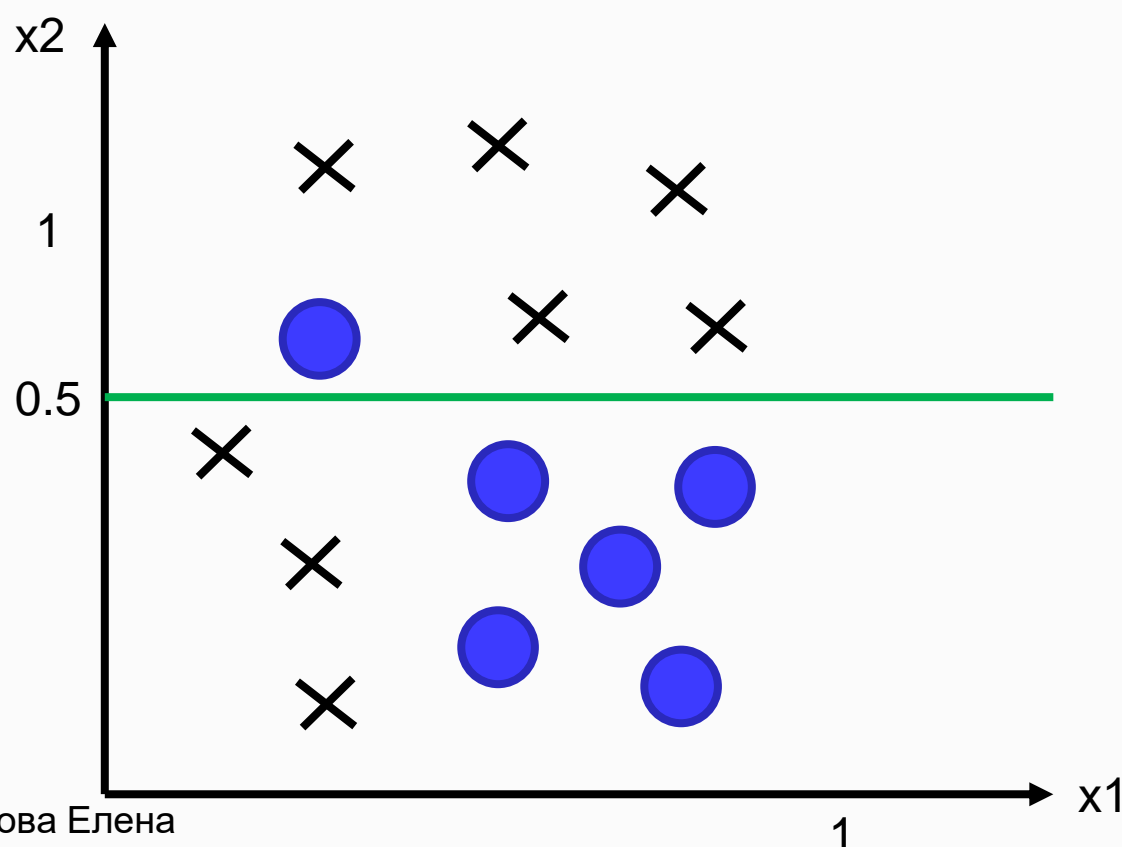
Пример: бинарная классификация

- Очевидно, эта задача классификации плохо решается линейной моделью

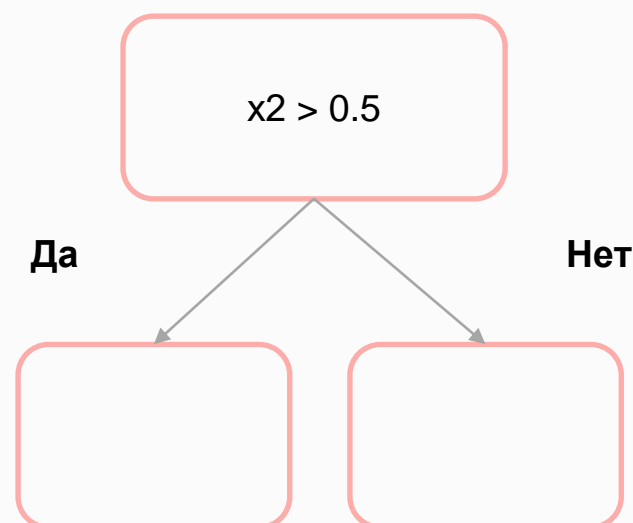
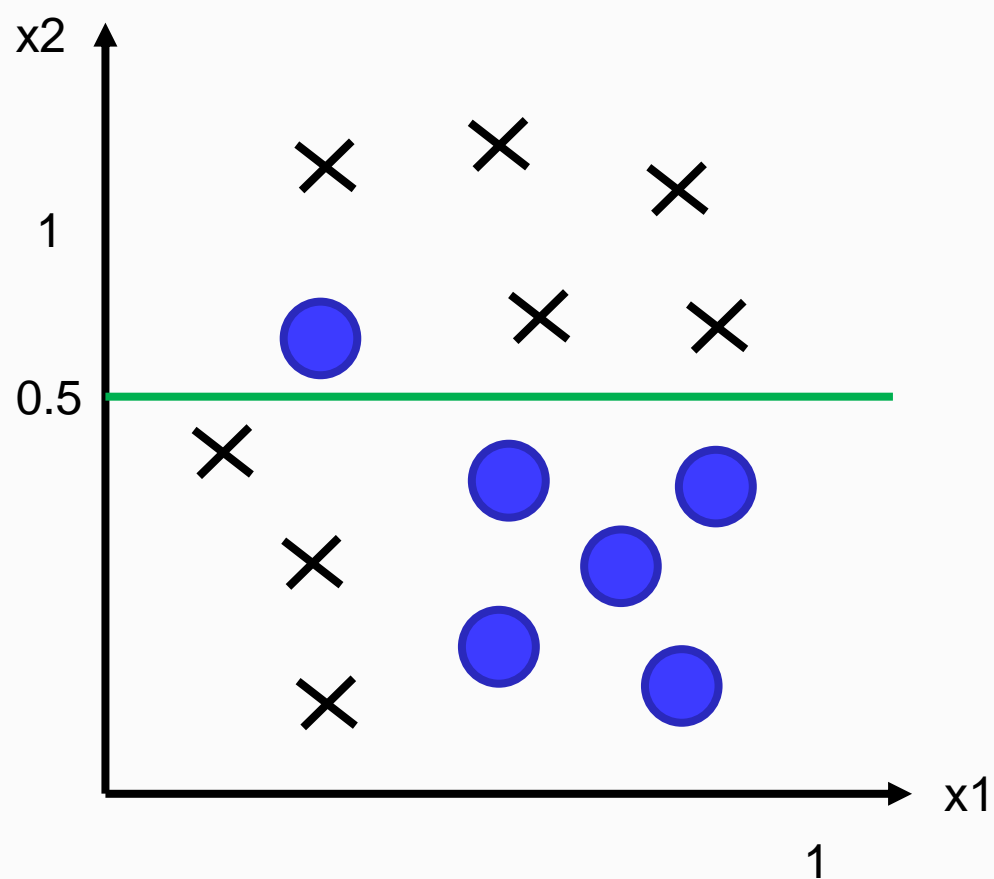


Жадный алгоритм

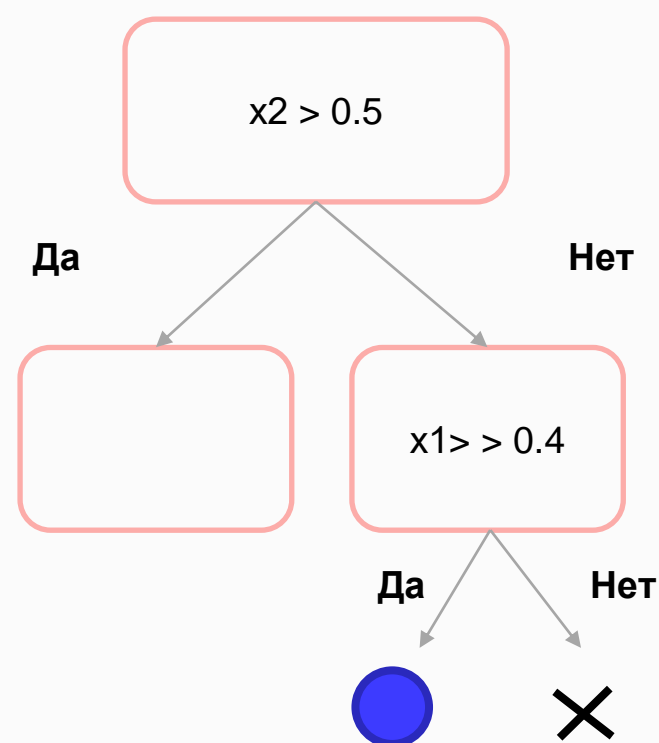
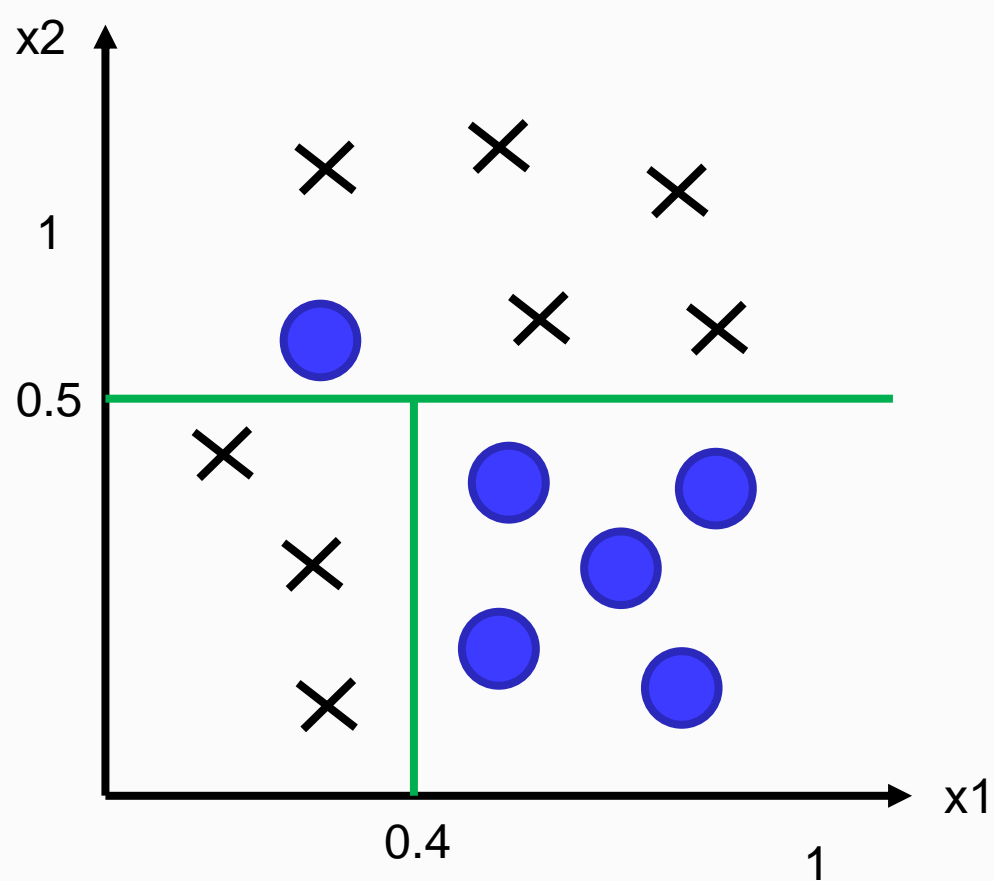
- Будем строить дерево жадным образом, то есть не сразу всё дерево, а по шагам: на каждом шаге будем выбирать предикат (или разбиение), дающий наименьшее число ошибок классификации



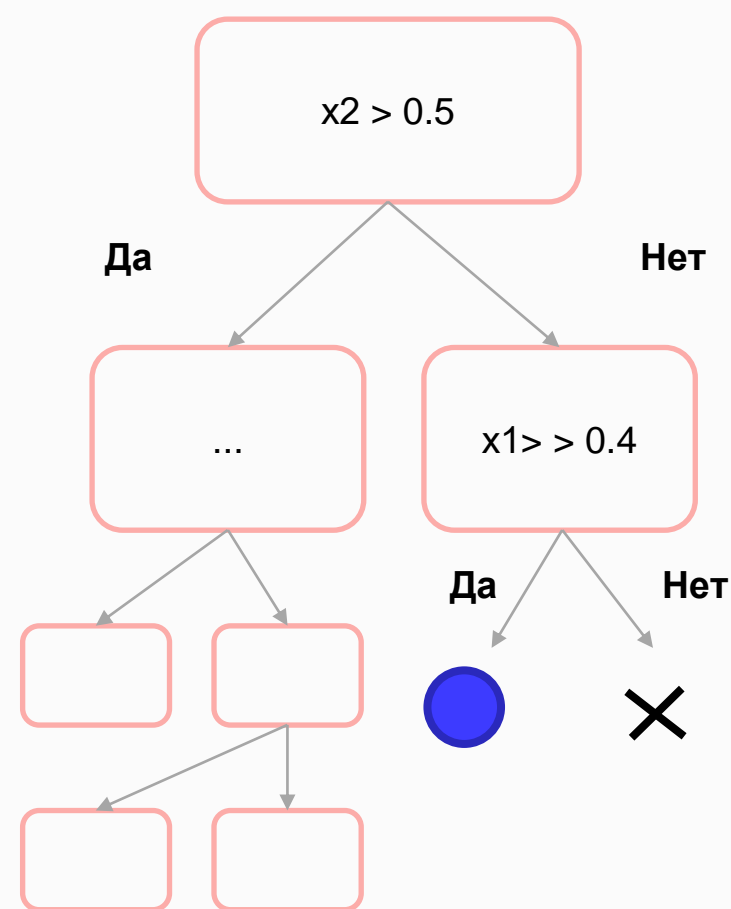
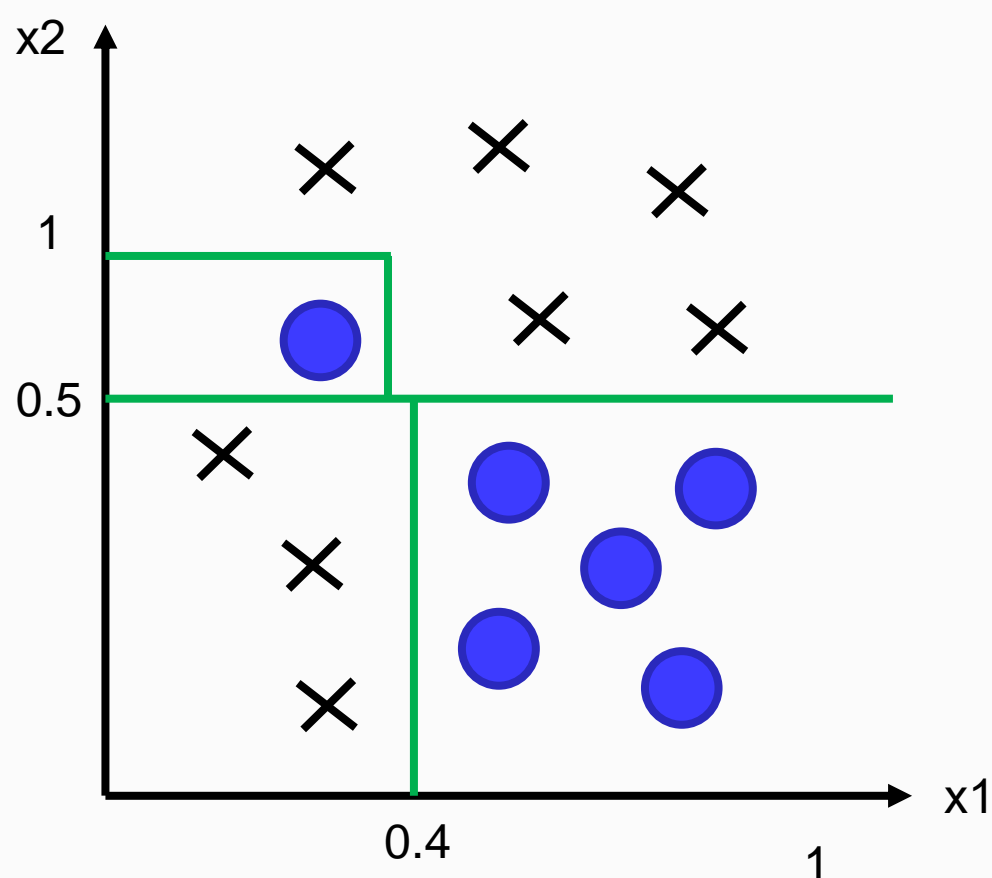
Жадный алгоритм: шаг 1



Жадный алгоритм: шаг 2

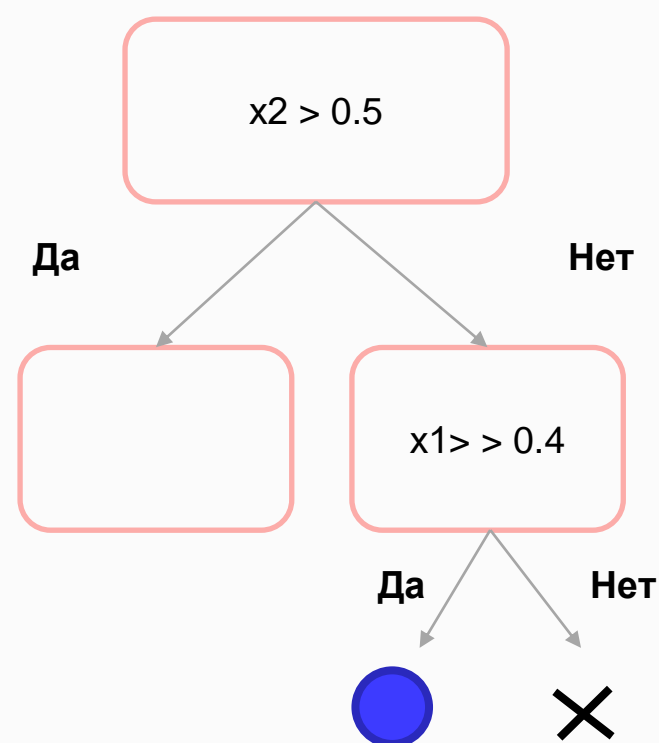
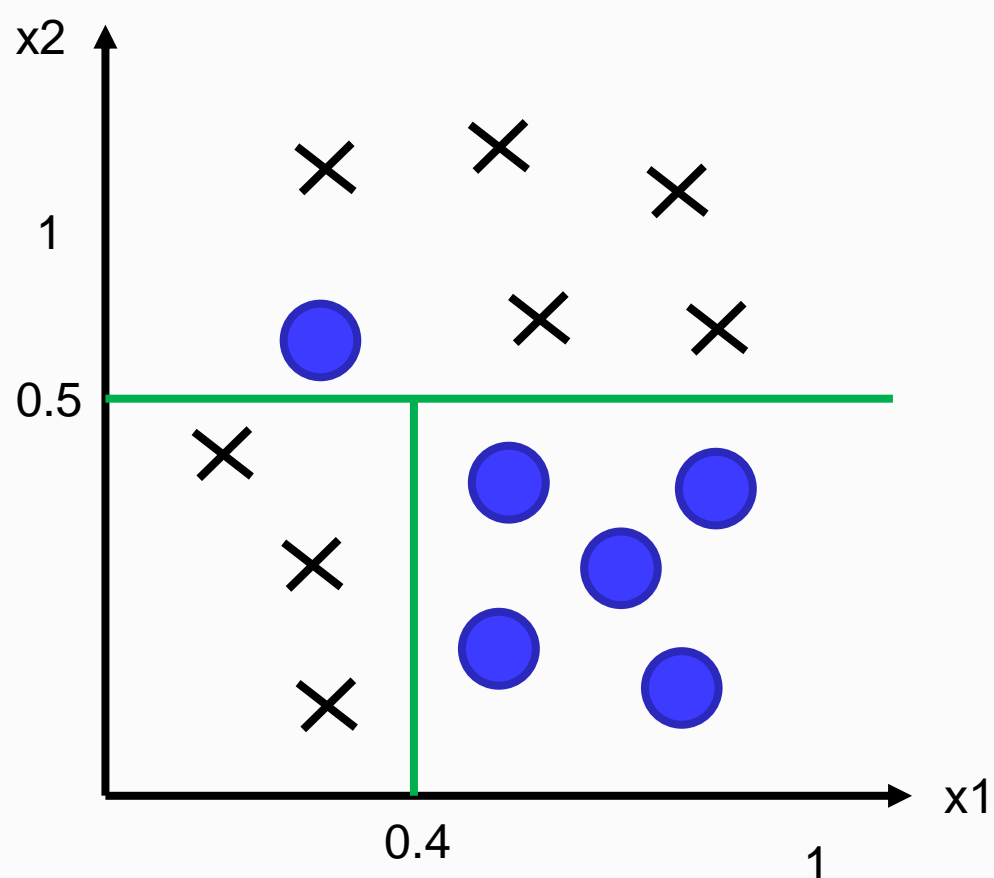


Жадный алгоритм: следующие шаги



- Получили переобученную модель

Лучше так



- Ошибка на обучении небольшая, но на тесте скорее всего будет лучше!

Важная особенность деревьев

- Важная особенность решающих деревьев — они склонны к сильному переобучению!
- Если не останавливать процедуру построения дерева, то в итоге дерево идеально подстроится под данные и в большинстве случаев гарантированно получит нулевую ошибку предсказания

Итоги

- ✓ Вы познакомились с терминологией, связанной с решающими деревьями
- ✓ Вы также посмотрели, как пошагово строится дерево (жадный алгоритм)