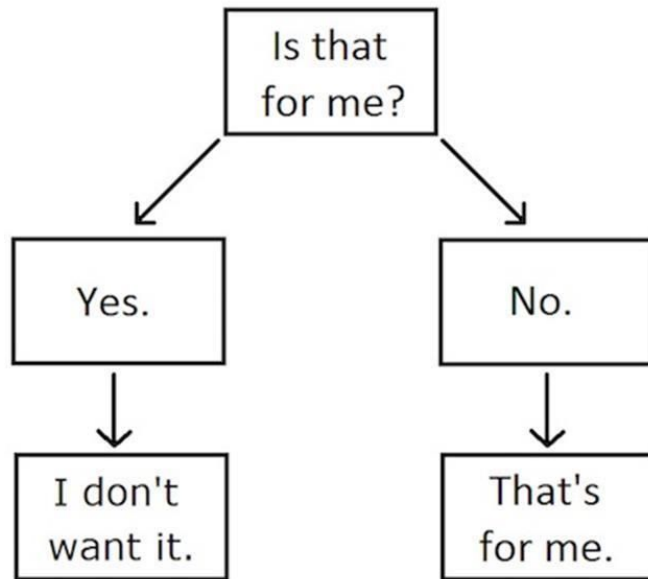


Обучение с учителем: Деревья решений (Decision Trees). Случайный лес (Random Forest).

Екатерина Кондратьева

My Cat's Decision-Making Tree.



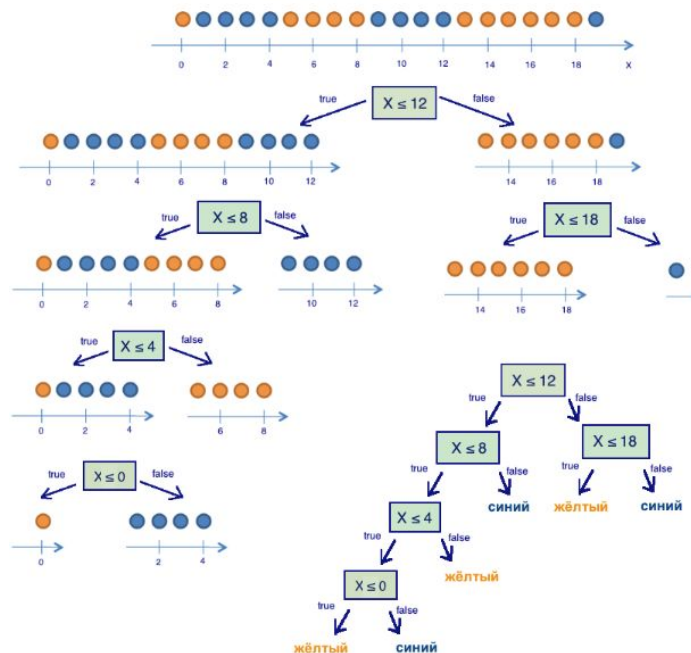
Деревья Решений

Дерево принятия решений (также может называться деревом классификации или регрессионным деревом) — средство поддержки принятия решений, использующееся в машинном обучении, анализе данных и статистике. Структура дерева представляет собой «листья» и «ветки». На рёбрах («ветках») дерева решения записаны атрибуты, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах — атрибуты, по которым различаются случаи. Чтобы классифицировать новый случай, надо спуститься по дереву до листа и выдать соответствующее значение.

Деревья принятия решений характеризуются:

- Критерием Информативности (в sklearn `gini`, `entropy`)
- Критерий Останова

Понятие энтропии. Принцип построения решающего правила в деревьях



Что означает
глубина дерева
(max depth)?

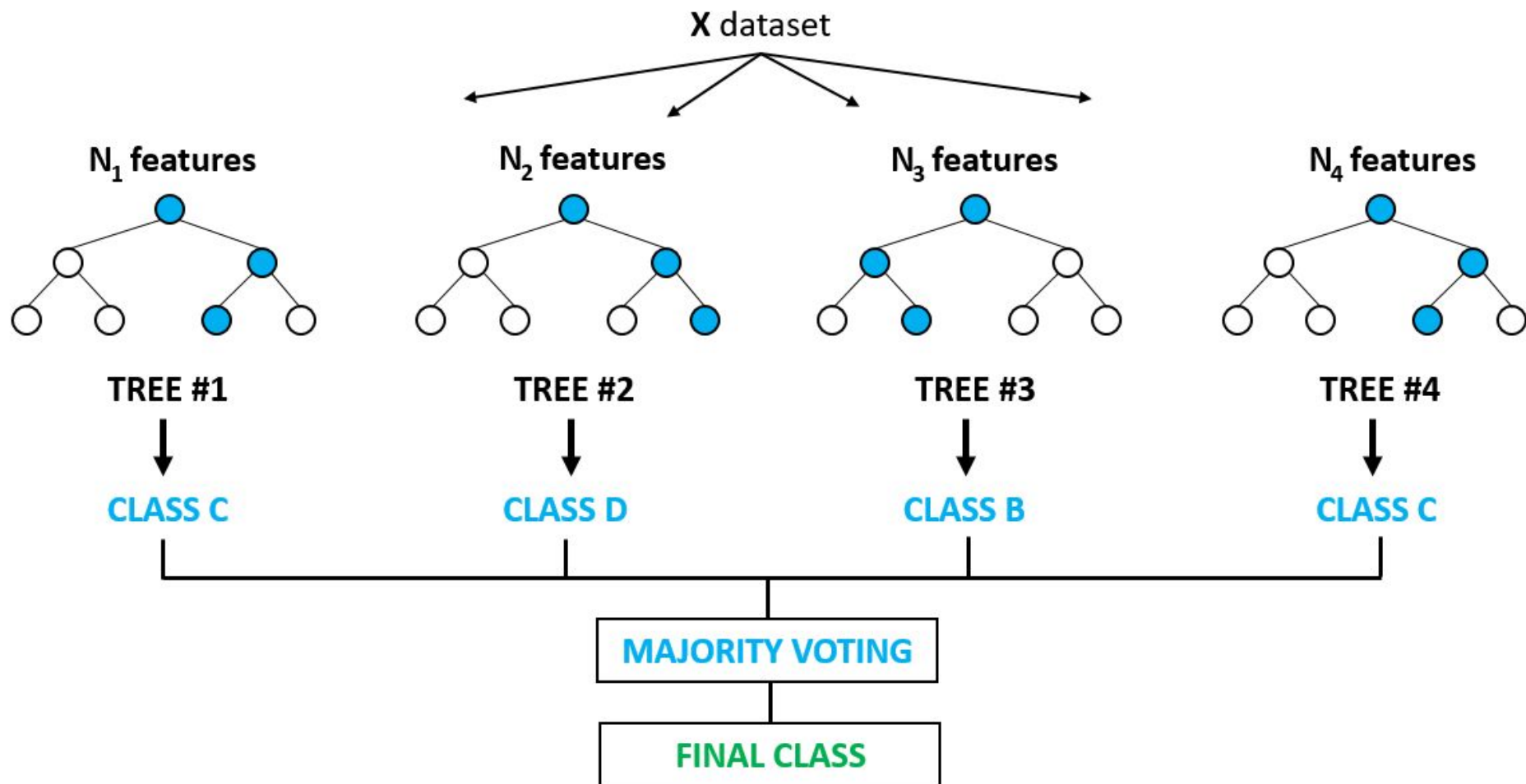


Случайные Леса Решений

Random forest (с англ. — «случайный лес») — алгоритм машинного обучения заключающийся в использовании комитета (ансамбля) решающих деревьев.

- Сгенерируем случайную подвыборку с повторениями размером N из обучающей выборки.
- Построим решающее дерево, классифицирующее образцы данной подвыборки, причём в ходе создания очередного узла дерева будем выбирать набор признаков, на основе которых производится разбиение (не из всех M признаков, а лишь из m случайно выбранных).
- Дерево строится до полного исчерпания подвыборки и не подвергается процедуре прунинга (англ. pruning — отсечение ветвей) (в отличие от решающих деревьев, построенных по таким алгоритмам, как CART или C4.5).

Алгоритм применяется для задач классификации, регрессии и кластеризации.



Вопросы для самопроверки:

1. В чем отличие Decision Trees от Random Forest?
2. На что влияют критерии построения решающего правила в деревьях?
3. Как интерпретировать результат модели RFC?
4. Почему важно варьировать `max_depth` дерева?

Источники:

1. Лекция <https://ru.coursera.org/lecture/supervised-learning/rieshaiushchiie-dieriev-ia-HZxD1>
2. https://chrisalbon.com/machine_learning/trees_and_forests/visualize_a_decision_tree/
3. <https://habr.com/ru/post/171759/>
4. <https://www.hse.ru/mirror/pubs/share/215285956>