

Деревья решений и ансамбли

https://classroom.github.com/a/_0S7sOh3

Набор данных

- Выберите любой набор данных для задачи классификации. Желательно использовать с первой лабораторной работы.
- Преобразуйте его в числовой вид. Допускается наличие нецелевых категориальных признаков.
- Разбейте его на тренировочную и тестовую часть.
- Выберите целевую функцию ошибки или качества.

Алгоритмы

- Реализуйте алгоритм построения дерева принятия решений. Алгоритм должен поддерживать не менее 3 гиперпараметров для ограничения размера дерева.
- Реализуйте алгоритм бустинга и случайного леса.

Задание

- Выберите библиотечную реализацию дерева принятия решений. Не ограничивая высоту, переберите по очереди разные значения числовых гиперпараметров. Постройте график зависимости высоты полученного дерева.
- Повторите предыдущий пункт для вашей реализации дерева принятия решений.
- Постройте график зависимости целевой функции ошибки или качества от высоты дерева на тренировочном и тестовом множестве для вашей и библиотечной реализации дерева принятия решений.
- Постройте график зависимости целевой функции ошибки или качества от числа деревьев на тренировочном и тестовом множестве для вашей и библиотечной реализации алгоритмов случайного леса.
- Преобразуйте набор данных, чтобы у него была бинарная целевая категория. Повторите предыдущий пункт для вашего и библиотечного алгоритма бустинга.