

Метод ближайших соседей

[github classroom](#)

Алгоритм

Реализуйте метод ближайших соседей:

- Алгоритм должен работать с окнами фиксированного и нефиксированного размера.
- Алгоритм должен работать с различными ядрами. Не менее 4 штук. Обязательно должно быть Равномерное и Гауссово ядро. Желательно, чтобы было Гауссово и общее ядро вида $(1 - |u|^a)^b$.
- Алгоритм должен работать с различными метриками. Не менее 3 штук. Обязательно должно быть Косинусное расстояние. Желательно, чтобы было Косинусное расстояние и расстояние Минковского L_p .
- Алгоритм должен работать с априорными весами.
- Разрешается использовать готовую реализацию алгоритма поиска ближайших объектов.
- Выберите аналогичную библиотечную реализацию метода ближайших соседей.

Набор данных

- Выберите любой набор данных для задачи классификации. Желательно использовать набор данных с курса Анализа данных.
- Преобразуйте его в числовой вид и нормализуйте.
- Разбейте его на тренировочную и тестовую часть.
- Выберите целевую функцию ошибки или качества.

Гиперпараметры

Найдите лучшие гиперпараметры:

- Гиперпараметры требуется настроить для реализованного и библиотечного алгоритма.
- Для поиска желательно использовать библиотеку `optuna`.
- Тестовое множество не должно использоваться для валидации при поиске.
- Выведите лучшие значения гиперпараметров для обоих алгоритмов.
- Для обоих алгоритмов постройте график зависимости целевой функции ошибки/качества на тестовом и тренировочном множестве в зависимости от числа соседей или ширины окна (смотря какая функция оказалась лучше). Остальные гиперпараметры должны быть зафиксированы.

Поиск аномалий

- Реализуйте алгоритм поиска аномалий LOWESS.
- Взвесьте объекты из тренировочного множества.
- Вычислите результат валидации реализованного алгоритма на тестовом множестве до и после взвешивания.
- Повторите предыдущий пункт с библиотечным алгоритмом. Если он не поддерживает априорные веса, используйте семплирование.