

## Домашнее задание по машинному обучению №1

Стоимость задач указана в скобках

Дедлайн на полный балл – 18 марта

Дедлайн на половинный балл – 25 марта

Датасет – `spambase.csv`

Класс – поле `label` (1 – спам, 0 – не спам)

1. **(2)** Реализуйте алгоритм kNN классификации по k ближайшим соседям, используя простое евклидовое расстояние.
2. **(0.5)** Вычислите значение метрики LOO для всех вариантов количества соседей (k) от 1 до 10.
3. **(2)** Реализуйте алгоритм RadiusNeighbours классификации по соседям, лежащим на расстоянии меньше r (радиус), используя простое евклидовое расстояние.
4. **(0.5)** Найдите лучший радиус с помощью бинарного поиска.
5. **(1)** Нормализуйте датасет так, чтобы все признаки лежали в отрезки [0, 1] и повторите вычисления, сделанные в пунктах 2 и 4, на новом датасете.

## Задачи на дополнительные баллы

Дедлайн на полный балл – 25 марта

Дедлайн на половинный балл – 1 апреля

1. **(1)** Сравните получившиеся результаты с результатами, полученные алгоритмом `KNeighborsClassifier` библиотеки `sklearn`.
2. **(1 или 2)** Реализуйте быстрый поиск ближайших соседей с помощью любой библиотеки построения `KDTree` (1) или напишите свою функцию поиска с помощью `KDTree`, дающую выигрыш по времени (2).

