

## Домашнее задание по машинному обучению №1

Стоимость задач указана в скобках

Дедлайн на полный балл – 15 марта

Дедлайн на половинный балл – 22 марта

Датасет – `spambase.csv`

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

1. **(4)** Реализуйте алгоритм kNN классификации по k ближайшим соседям, используя простое евклидовое расстояние.
2. **(1)** Вычислите значение метрики LOO для всех вариантов количества соседей (k) от 1 до 10.
3. **(4)** Реализуйте алгоритм RadiusNeighbours классификации по соседям, лежащим на расстоянии меньше r (радиус).
4. **(1)** Найдите лучший радиус с помощью тернарного поиска.
5. **(2)** 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 – 22 марта

Дедлайн на половинный балл – 29 марта

1. **(4)** Реализуйте алгоритм WkNN с весом  $\max[0, \frac{r - \rho(x_i, x)}{r}]$ , и подберите лучшую константу r.
2. **(4)** Реализуйте алгоритм KDTree и сравните время работы.