Домашнее задание по машинному обучению №1

Стоимость задач указана в скобках

Дедлайн на полный балл – **27 февраля**

Дедлайн на половинный балл – **9 марта**

**К 6 марта должно быть минимум 6 баллов**

Датасеты – spam.csv, cancer.csv

1. (3) Реализуйте алгоритм kNN по k ближайшим соседям, используя простое евклидовое расстояние.

2. (2) Вычислите значение метрики LOO вашего алгоритма на обоих датасетах для всех вариантов количества соседей (k) от 1 до 10.

Подсказка: для более быстрого вычисления метрики можно для каждой точки сначала выписать список 11-ти ближайших соседей в порядке увеличения расстояния. Потом можно выкинуть ближайшую точку (это будет сама целевая точка) и взять нужное количество соседей.

3. (3) Реализуйте алгоритм RadiusNeighbours по соседям, лежащим на

расстоянии меньшем r.

4. (2) Найдите лучший радиус по LOO.

Подсказка: можно использовать аналогичный пункту 2 подход.

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.

Дополнительные задачи:

1. (0) Реализуйте алгоритм WkNN с весом max[0, (r−ρ(xi,x))/r], и подберите лучшую константу r.