

Лабораторная работа 2.

Тема: Метод k-ближайших соседей

Цель работы: разработка классификатора на основе метода k ближайших соседей.

Задание

1. Загрузите набор данных в соответствии с вариантом:
 $N_{\phi} = ((N_c + 2) \bmod 5) + 1$, где N_{ϕ} – номер файла, N_c – номер по списку группы.
2. Данные из файла необходимо разбить на две выборки, обучающую и тестовую, согласно общепринятым правилам разбиения. (70/30, 80/20). Разбиение выборки необходимо выполнять программно, случайным образом, при этом, не нарушая информативности обучающей выборки. Разбивать рекомендуется по следующему правилу: делим выборку на 3 равных части, 2 части используем в качестве обучающей, одну в качестве тестовой. Проанализируйте обучающую выборку на возможность минимизировать разницу между количеством представленных в ней объектов разных классов.
3. На основе этих данных необходимо обучить разработанный классификатор, отразив метод подбора параметров в соответствии с вариантом (k, i, q, h), и протестировать метод на тестовой выборке. Вариант алгоритма выбирается следующим образом:

$N_b = (N_c \bmod 3) + 1$, где N_b – номер варианта, N_c – номер по списку группы

1). Метод k взвешенных ближайших соседей

Для этого варианта необходимо использовать весовую функцию w_i по формуле $N_b = (N_c \bmod 2) + 1$. Параметр q подбирается методом скользящего контроля.

1.) $w_i = q^i, q \in (0,1)$

2.) $w_i = \left(\frac{k+1-i}{k}\right)^q, q \in (2,3,4)$

2). Метод парзеновского окна с фиксированным h

В данном варианте необходимо использовать функцию ядра $K(z)$, выбранную следующим образом:

$$N_{\text{я}} = (N_c * 6 + 13) \bmod 8 \bmod 3 + 1$$

1. Q –квартичное $K(x) = (1 - r^2)^2[r \leq 1]$
2. T –треугольное $K(x) = (1 - r)[r \leq 1]$
3. P –прямоугольное $K(x) = [r \leq 1]$

3). Метод парзеновского окна с относительным размером окна

В данном варианте аналогично необходимо использовать функцию ядра $K(z)$, выбранную следующим образом:

$$N_{\text{я}} = (N_c * 6 + 13) \bmod 8 \bmod 3 + 1$$

1. Q –квартичное $K(x) = (1 - r^2)^2[r \leq 1]$
2. T –треугольное $K(x) = (1 - r)[r \leq 1]$
3. P –прямоугольное $K(x) = [r \leq 1]$

4. В качестве отчёта требуется представить работающую программу и таблицу с результатами тестирования для каждого из 10 разбиений.
5. Оформите отчет в электронном виде, приложив ссылку на Jupyter Notebook/ Google Colab, где код сопровождается краткими выводами по каждому шагу, электронный вид отчёта в формате **pdf** в ЭИОС.

Дополнительная информация о задании:

- Как правило, значения параметров классификатора (например, количество соседей k) по критерию скользящего контроля с исключением объектов по одному (**Leave-one-out**). Для каждого объекта $x_i \in X_i$ проверяется, правильно ли он классифицируется по своим k ближайшим соседям.
- При использовании весовой функции возможно совпадение суммарного веса для нескольких классов. Это приводит к неоднозначности ответа при классификации. Чтобы такого не происходило, используют функцию ядра $K(z)$.
- При **фиксированной** ширине окна:
 - величина h фиксируется с заданным числовым значением; все объекты, которые попали в окно радиусом h участвуют в голосовании.
- При **относительной** ширине окна:
 - величина h , подаваемая в функция нахождения веса, берется от самого дальнего k -го соседа.