

Министерство цифрового развития

Федеральное государственное бюджетное образовательное учреждение высшего
образования

«Сибирский государственный университет телекоммуникаций и
информатики»
(СибГУТИ)

Кафедра прикладной математики и кибернетики

Отчёт

по лабораторной работе № 2 «Метод k ближайших соседей»

Выполнил:

студент группы ИП-213

Дмитриев Антон Александрович

Работу проверил: Преподаватель

Сороковых Дарья Анатольевна

Новосибирск 2025 г.

Цель работы

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

Задание

1. Загрузите набор данных в соответствии с вариантом:
 $N\phi = ((Nc + 2) \bmod 5) + 1$, где $N\phi$ – номер файла, Nc – 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:

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

1). Метод k взвешенных ближайших соседей

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

- 1.) $w_i = q^i, q \in (0,1)$
- 2.) $w_i = (k+1-i)q, q \in (2,3,4)$

2). Метод парzenовского окна с фиксированным h

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

$$Nя = (Nc * 6 + 13) \bmod 8 \bmod 3 + 1$$

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

3). Метод парzenовского окна с относительным размером окна

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

$$N\mathbf{я} = (Nc * 6 + 13) \bmod 8 \bmod 3) + 1$$

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

Основная часть

Номер в журнале 5

Вариант 3 $((5 \bmod 3) + 1 = 3)$ Метод парзеновского окна с относительным размером окна

Вариант 1 $((5 * 6 + 13) \bmod 8) \bmod 3 + 1 = 1$ Функция квадратического ядра

Ссылка на google colab:

<https://colab.research.google.com/drive/19G6oV5s8Fiu4ni-EmBEOuCSr8Lyp4Sx4?usp=sharing>

Таблица с результатами тестирования для каждого из 10 разбиений:

Итоговые результаты для 10 разбиений:

	split	best_k	accuracy	train_size	test_size
0	1	3	0.851667	7000	3000
1	2	3	0.861667	7000	3000
2	3	3	0.844667	7000	3000
3	4	3	0.854333	7000	3000
4	5	3	0.847667	7000	3000
5	6	3	0.849333	7000	3000
6	7	3	0.860333	7000	3000
7	8	3	0.837000	7000	3000
8	9	3	0.848333	7000	3000
9	10	3	0.848667	7000	3000

Средняя точность: 0.8504 ± 0.0072