НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

**Лабораторна робота № 4**

з дисципліни «Прикладні задачі машинного навчання»

Тема: «Класифікація методом k найближчих сусідів і набір даних Digits»

|  |  |
| --- | --- |
| **Прийняв:** | **Виконав:**  **студент групи ІП-13**  **Недельчев Є.О.** |

Київ – 2023

**Завдання:**

**1. Для дослідження даних, візуалізуйте їх. Виведіть зображення перших 36 цифр з набору**

**2. Розбийте дані на навчальні та тестові, за замовчуванням train\_test\_split резервує 75% даних для навчання і 25% для тестування, змініть це.**

**3. Створити та навчити модель**

**4. Виконайте прогнозування класів**

**5. Порівняйте прогнозовані цифри з очікуваними для перших 36 тестових зразків.**

**6. Поясніть результат, застосуйте метрики точності моделі.**

**7. Виведіть звіт класифікації**

**8. Використайте декілька моделей KNeighborsClassifier, SVC и GaussianNB для пошуку найкращої**

**9. Налаштуйте гіперпараметр K в KNeighborsClassifier**

**Виконання**

**1. Для дослідження даних, візуалізуйте їх. Виведіть зображення перших 36 цифр з набору**

Імпортуємо необхідні бібліотеки:

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Завантажимо датасет та візуалізуємо його за допомогою бібліотеки matplotlib.

**Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание**

**2. Розбийте дані на навчальні та тестові, за замовчуванням train\_test\_split резервує 75% даних для навчання і 25% для тестування, змініть це.**

**Изображение выглядит как текст, линия, Шрифт, снимок экрана

Автоматически созданное описание**

**3. Створити та навчити модель.**

**Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание**

**4. Виконайте прогнозування класів**

**Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание**

**5. Порівняйте прогнозовані цифри з очікуваними для перших 36 тестових зразків.**

**Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание**

Як можна помітити, жодних розбіжностей між очікуваними та прогнозованими даними не виявлено.

**6. Поясніть результат, застосуйте метрики точності моделі.**

Спочатку оцінимо якість моделі за допомогою вбудованого методу score:

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Створимо матрицю невідповідностей та виведемо її:

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Як можна побачити, модель є дуже точною, помилкові результати трапляються, але абсолютна більшість прогнозів є вірною.

**7. Виведіть звіт класифікації**

Для цього скористаємося функцією з модулю sklearn.metrics:

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

**8. Використайте декілька моделей KNeighborsClassifier, SVC и GaussianNB для пошуку найкращої**

Спочатку створим та навчимо моделі SVC та GaussianNB:

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Порівняємо їхню точність за допомогою вбудованого метода score:

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Можна зробити висновок, що метод опорних векторів показав себе найточнішою моделлю.

**9. Налаштуйте гіперпараметр K в KNeighborsClassifier**

Поекспериментуємо з гіперпараметром k для пошуку найкращого значення:

**Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание**

Найкращим виявився параметр k=3 (за замовчуванням він дорівнює 5).

**Висновок**

Виконуючи цю лабораторну роботу я навчився реалізовувати основні етапи машинного навчання, а саме:

* Вибір даних для навчання моделі.
* Завантаження та аналіз даних.
* Розбиття даних для навчання і тестування.
* Вибір і побудова моделі.
* Навчання моделі.
* Формування прогнозів.

Я провів оцінку результатів та наочно їх відобразив у роботі. Поекспериментував з гіперпараметрами моделей і знайшов найкращий. Також було опрацьовано декілька класифікаційних моделей для пошуку найкращої серед них.