НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

**Лабораторна робота № 3**

з дисципліни «Прикладні задачі машинного навчання»

Тема: «Класифікація, регресія і кластеризація з використанням бібліотеки scikit-learn»

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

Київ – 2023

**Завдання:**

**1. Повторити дії описані в пункті «Часові ряди і проста лінійна регресія частина 2» даної лабораторної роботи та порівняти з результатом попередньої лабораторної роботи.**

**2. Аналогічно з прикладом з лекції 7 згенеруйте набір даних та класифікуйте його використавши класифікатор SVC (слайд 95).**

**3. Порівняти декілька класифікаційних оцінювачів наприклад KNeighborsClassifier, SVC та GaussianNB для вбудованого в scikit-learn одного набору даних (вибрати довільний за бажанням).**

**Виконання**

**1. Повторити дії описані в пункті «Часові ряди і проста лінійна регресія частина 2» даної лабораторної роботи та порівняти з результатом попередньої лабораторної роботи.**

1. Імпортуємо необхідні бібліотеки

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

2. Завантажимо дані, перейменуємо стовпець ‘Value’ в ‘Temperature’ та видалимо 01 в кінці кожного значення дати і виведемо кілка зразків даних

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

3. Розбиваємо дані на навчальний і тестовий набори. Перевіримо пропорції навчальних і тестових даних (75% і 25%).

Изображение выглядит как текст, Шрифт, линия, снимок экрана

Автоматически созданное описание

4. Скористаємося оцінювачем LinearRegression. Виведемо значення кута нахилу і точки перетину прямої з віссю.

Изображение выглядит как текст, снимок экрана, Шрифт, число

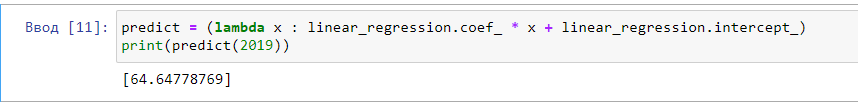
Автоматически созданное описание

5. Виведемо результати моделі та порівняємо з реальними.

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

6. Спрогнозуємо результат для 2019 року.



В минулій лабораторній роботі за допомогою модулю stats було отримано результат 64.46, що відхиляється від реального значення на 1.68. В цій лабораторній роботі також була використана лінійна регресія, але вже з модуля sklearn, і як можна побачити, результат від попереднього відрізняється зовсім не суттєво і також несе в собі достатньо серйозну погрішність.

7. Побудуємо діаграму розкиду даних.

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Автоматически созданное описание

**2. Аналогічно з прикладом з лекції 7 згенеруйте набір даних та класифікуйте його використавши класифікатор SVC (слайд 95).**

1. Згенеруємо дані (двовимірний масив з 200 зразків і 2 ознаками). Створимо мітки для даних. Спочатку перевіримо, чи значення першої ознаки в кожному зразку більше за 0, а потім перевіримо, чи значення другої ознаки в кожному зразку більше за 0. Виконаємо логічну операцію XOR між цими двома умовами, щоб отримати кінцеву мітку. Якщо задовольняється тільки одна з двох умов, мітка дорівнює 1, в іншому випадку - мітка дорівнює -1. Візуалізуємо отримані дані:

Изображение выглядит как текст, снимок экрана, Красочность, диаграмма

Автоматически созданное описание

2. Навчимо модель на наших даних , використовуючи алгоритм SVM з ядром rbf. Потім візуалізуємо прогнози моделі на тестових даних у вигляді контурного графіка: Изображение выглядит как текст, снимок экрана, карта, диаграмма

Автоматически созданное описание

**3. Порівняти декілька класифікаційних оцінювачів наприклад KNeighborsClassifier, SVC та GaussianNB для вбудованого в scikit-learn одного набору даних (вибрати довільний за бажанням).**

1. Було обрано набір даних про ракові захворювання. Завантажимо ці дані та виведемо перші 5 для кращого їх розуміння.

Изображение выглядит как текст, число, Шрифт, программное обеспечение

Автоматически созданное описание

2. Створимо три моделі класифікації (k-найближчих сусідів (KNN), з параметром n\_neighbors=5, метод опорних векторів SVM з лінійним ядром та наївний байєсівський класифікатор GaussianNB) та навчимо їх на тренувальних даних. Изображение выглядит как текст, снимок экрана, дисплей, Шрифт

Автоматически созданное описание

3. Обчислимо точність цих класифікаторів та візуалізуємо її для кращого порівняння.

Изображение выглядит как текст, снимок экрана, дизайн

Автоматически созданное описание

Як можна побачити, найбільшу точність продемонстрував метод опорних векторів SVC.

**Висновок**

Виконуючи цю лабораторну роботу я ознайомився з бібліотекою seaborn та використав на практиці графік лінійної регресії. Також були спрогнозовані дані на підставі старих даних та було зроблено висновок, що прогнозування подій із використанням лінійної регресії є досить неточним і дозволяє лише приблизно оцінити необхідні нам дані.