**Міністерство освіти і науки України**

**Національний технічний університет України «Київський політехнічний**

**інститут імені Ігоря Сікорського"**

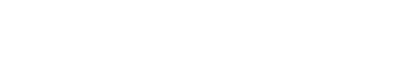
**Факультет інформатики та обчислювальної техніки**

**Кафедра інформатики та програмної інженерії**

# Звіт

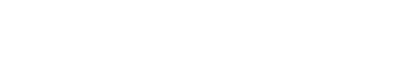
з лабораторної роботи № 2 з дисципліни

«Часові ряди і прості лінійна регресія»



**Виконав** *ІП-15, Дацьо Іван*

(шифр, прізвище, ім'я, по батькові)



**Перевірив**



прізвище, ім'я, по батькові



*Нестерук А.*

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

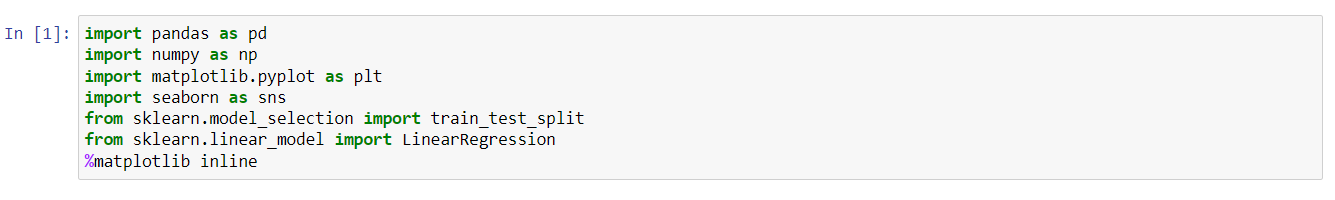
класифікуйте його використавши класифікатор SVC (слайд 95).

3. Порівняти декілька класифікаційних оцінювачів наприклад

KNeighborsClassifier, SVC та GaussianNB для вбудованого в scikit-learn одного набору даних (вибрати довільний за бажанням).

# Виконання

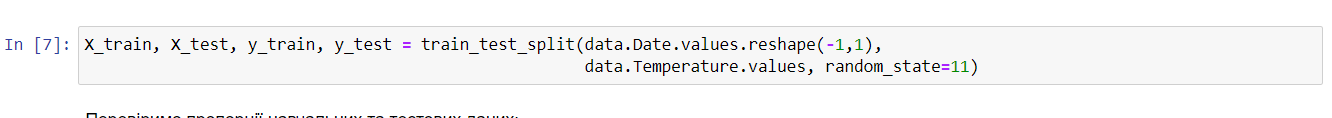
1. Імпортуємо необхідні пакети:



Завантажимо дані із csv файлу в dataframe та видалимо не потрібну інформацію місяця із колонки Date:

****

Розбиваємо дані на навчальний і тестовий набори:

****

Перевіримо пропорції навчальних та тестових даних(75% до 25%):

****

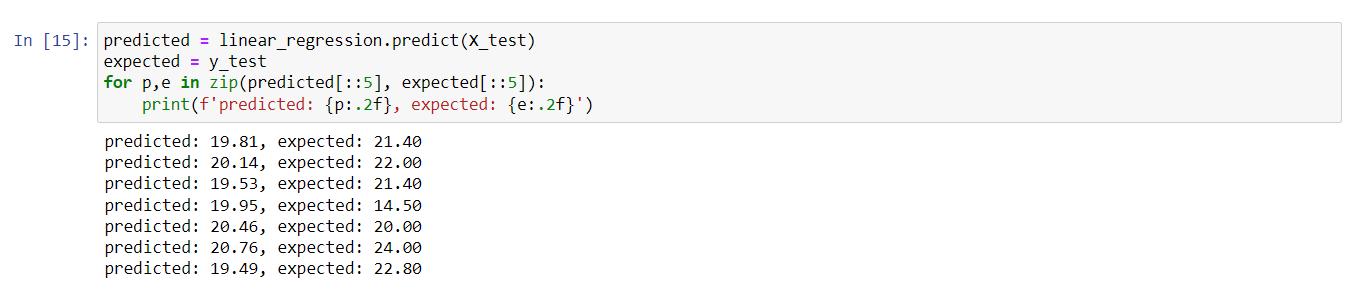
Використаємо оцінювач LinearRegression:



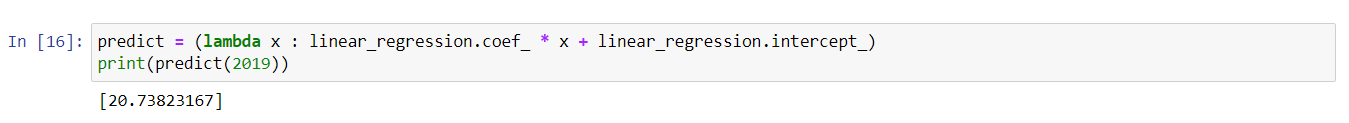
Виведемо значення кута нахилу та точку перетину:



Виведемо прогнозовані і очікувані значення для кожного п'ятого елементу:

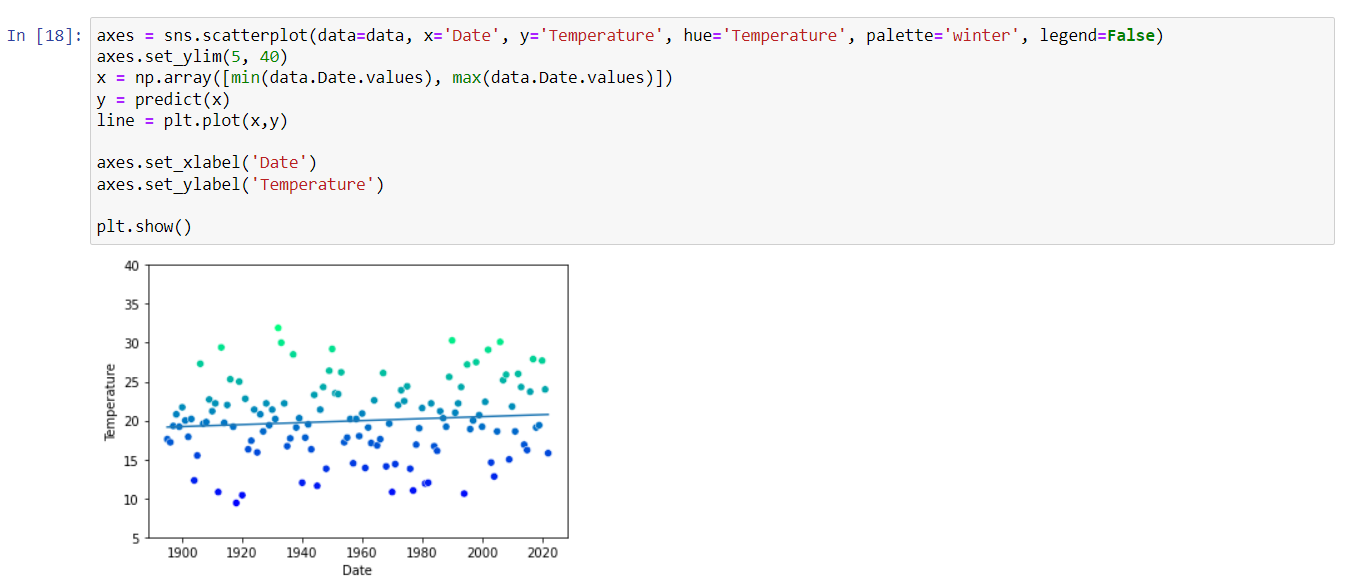


Спрогнозуємо дані для 2019 року:



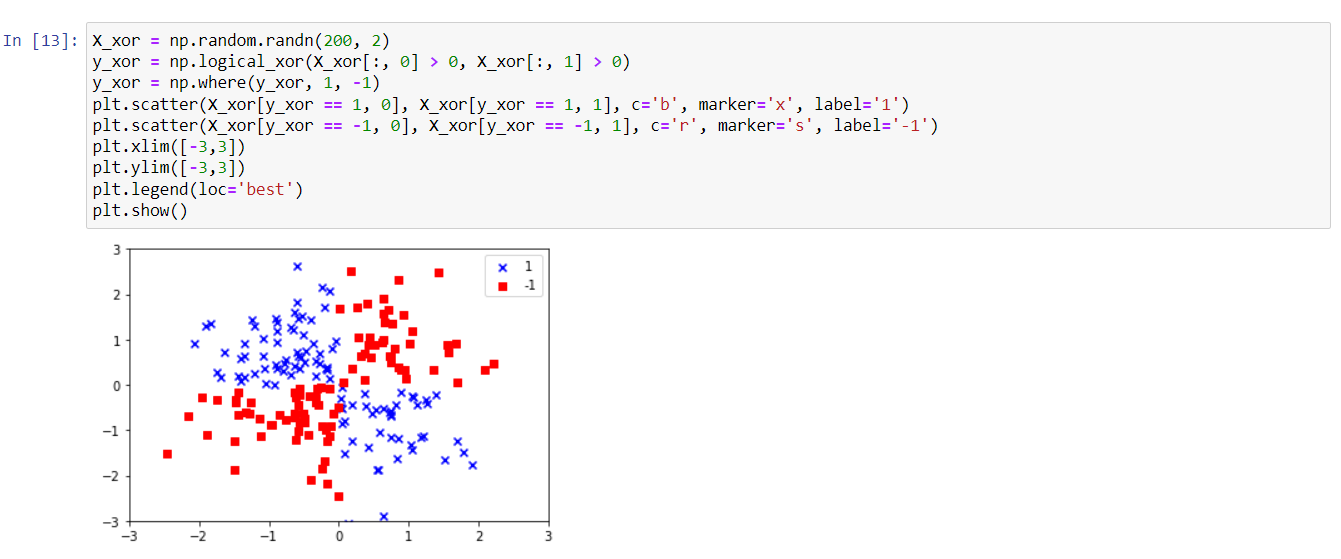
В минулій лабораторній використовуючи модуль stats було отримано результат для 2019 року 20.90, що відрізнявся від реального на +1.5. В даній лабораторній роботі використовуючи модуль sklearn і отримали 20.73, який відрізняється від реально на +1.3. Тобто даний метод також має значну погрішність.

Виведемо діаграму розкиду даних:

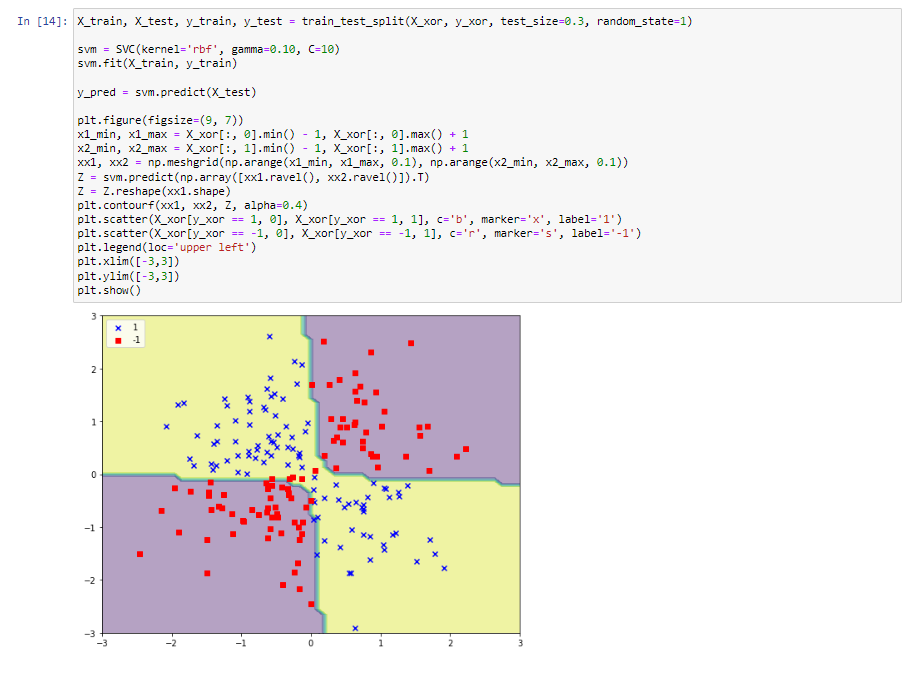


1. Імпортуємо додаткові пакети: 

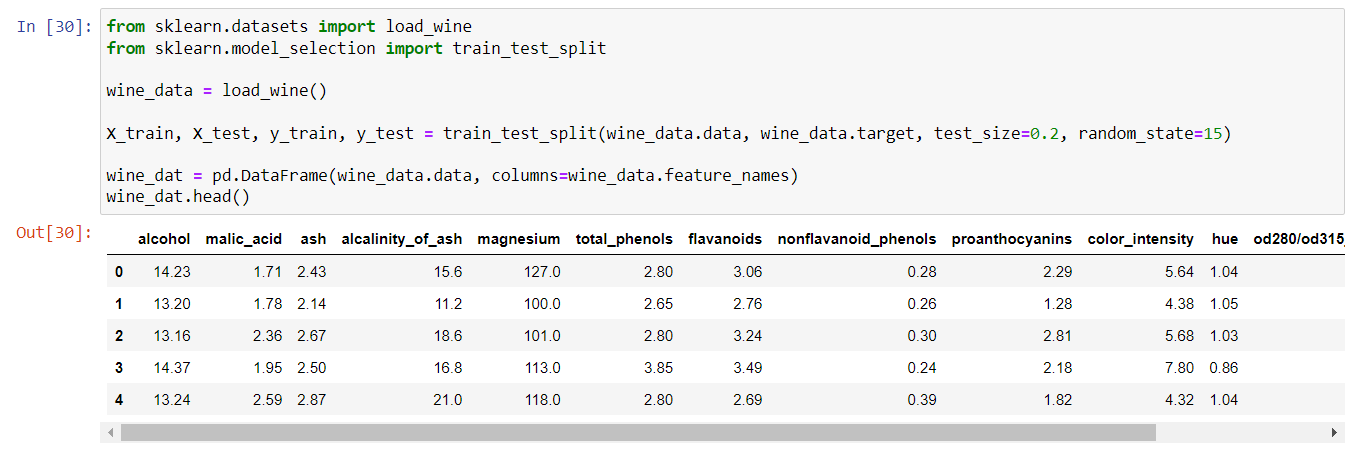
Згенеруємо масив із 200 записами за 2 ознаками. Для цих даних створимо мітки. Перевіримо значення ознаки у зразках на додатність. Виконаємо логічну операцію ХОR між умовами для отримання кінцевої мітки. Візуалізуємо також дані:



Виконаємо навчання на даних використовуючи алгоритм SMV та ядро rbf. Виконаємо візуалізацію використовуючи контурний графік:

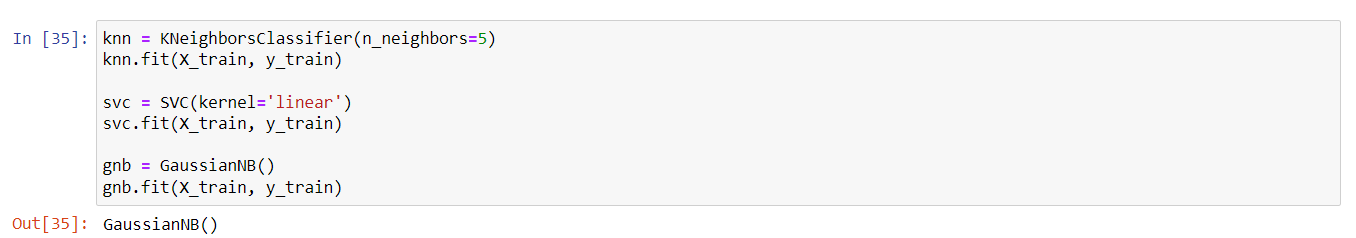


1. Завантажимо набір даних про вино load\_wine та виведемо елементи у шапці:



Завантажуємо додаткові пакети:

Створимо три класифікаційних оцінювачів, а саме KNeighborsClassifier, SVC(з лінійним ядром), та GaussianNB. Виконаємо навчання на тренувальних даних:



Оцінимо точність класифікаторів та побудуємо для них гістограму для наочної візуалізації:



**Висновок**

Виконавши дану лабораторну я ознайомився з бібліотекою scikit-learn та досліджено прогнозну здатність простої лінійної регресії на часових рядах, виконано класифікацію за допомогою класифікатора SVC і порівняно декілька класифікаційних оцінювачів. Можна зробити висновок що прогнозування використовуючи лінійну регресію є не точним методом і він дає лише приблизні значення, яких достатньо для загальної оцінки.