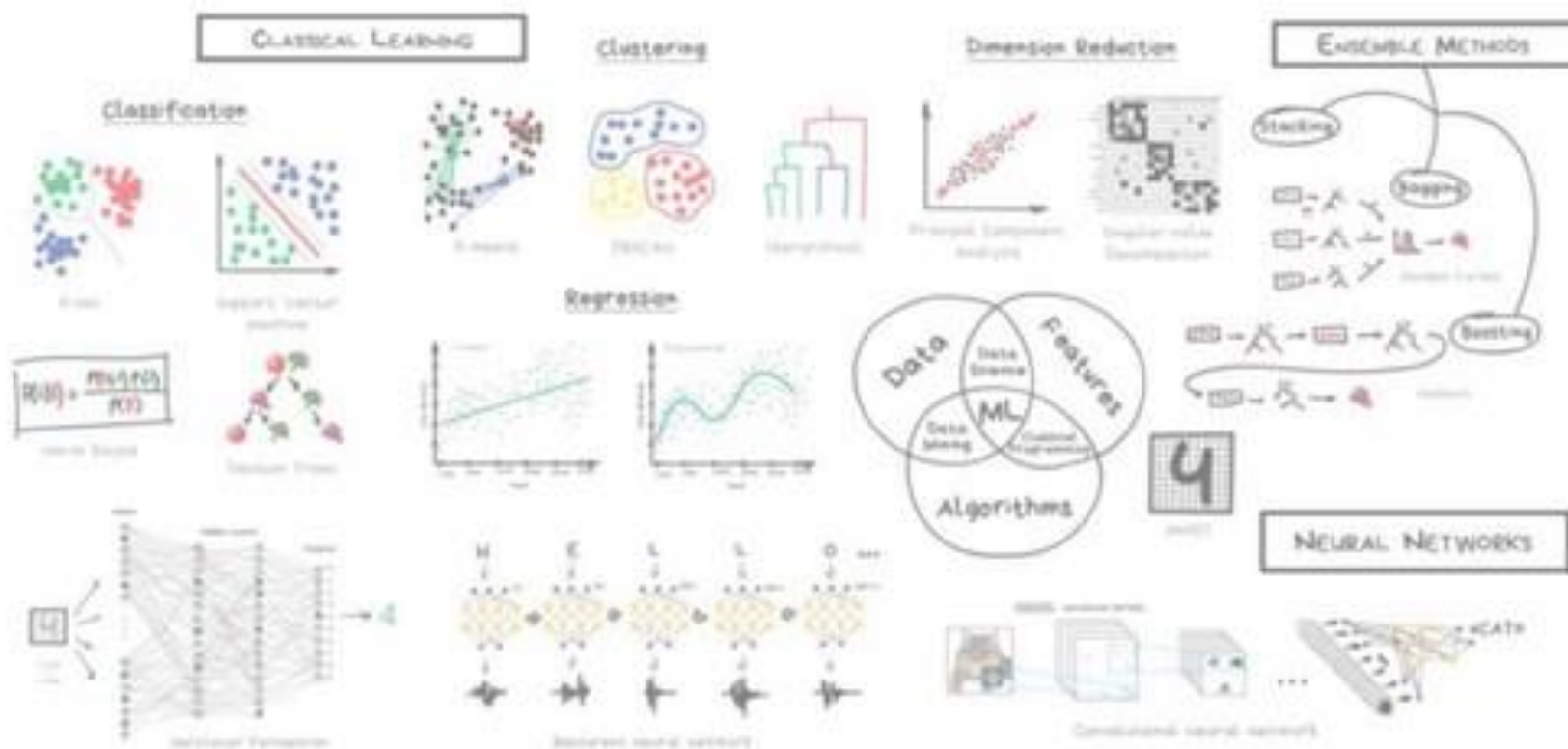


МАШИННЕ НАВЧАННЯ

Метод опорних векторів (SVM)

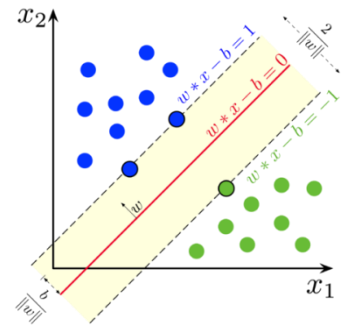


Лабораторна робота №3

Метод опорних векторів (SVM)

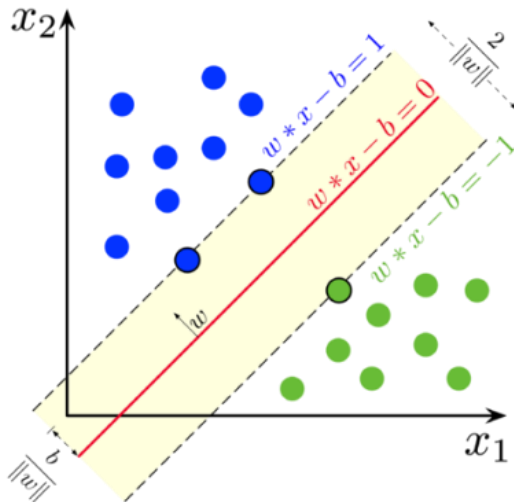
Постановка задачі

Розглянемо задачу бінарної класифікації, в якій об'єктам з $X = \mathbb{R}^2$ (об'єкти описуються **двома** числовими признаками): $X = \{x_1, x_2\}$ відповідає один з двох класів $Y = \{-1, +1\}$. Нехай задана вибірка пар "об'єкт-відповідь": $(x_i, y_i), i = 1 \dots N$. Необхідно побудувати алгоритм класифікації $a(x): X \rightarrow Y$.



Етапи розв'язання

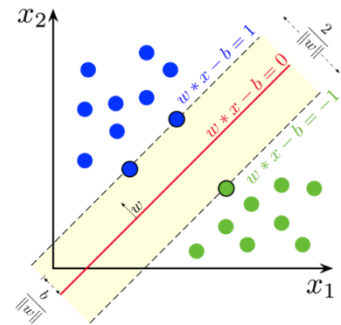
Розділяюча полоса



1. Провести первинну обробку даних вибірки:
 - Визначити дві ознаки $X = \{x_1, x_2\}$
 - Визначити два класи, з умовою $Y = \{-1, +1\}$
2. Розділити всю вибірку на навчальну та тестову
3. Побудувати алгоритм навчання на навчальній вибірці
4. Перевірити точність роботи алгоритму на тестовій вибірці
5. Порівняти результати з SVM з sklearn
6. Оформити результати у вигляді звіту.

Метод опорних векторів (SVM) в задачах класифікації

1. Первинна обробка даних



$$x = (x_1, x_2, \dots, x_n)$$

	A	B	C	D	E
1	Radius_mean	Texture_mean	Perimeter_mean	Area_mean	Diagnosis
2	17.99	10.38	122.8	100	1
3	20.57	17.77	132.9	132	1
4	19.69	21.25	130	120	1
5	11.42	20.38	77.58	186	1
6	20.29	14.34	135.1	129	1
7	12.45	15.7	82.57	477	1
8	18.25	19.98	119.6	104	1
9	13.71	20.83	90.2	577	1
10	13	21.82	87.5	519	1
11	12.46	24.04	83.97	475	1
12	16.02	23.24	102.7	797	1
13	15.78	17.89	103.6	78	1
14	14.61	15.69	92.68	664	0
15	12.76	13.37	82.29	504	0
16	11.54	10.72	73.73	409	0
17	8.597	18.6	54.09	221	0
18	12.49	16.85	79.19	481	0
19	12.18	14.08	77.25	461	0
20	18.22	18.87	118.7	102	1
21	9.042	18.9	60.01	244	0

$$x_{n+1} = 1; w_0 = -b$$

$$[\langle x_i, w \rangle - b]_{i=1 \dots n} \rightarrow [\langle x_i, w \rangle]_{i=1 \dots 3}$$

- 1.1. Визначити 2 головних признаки за якими буде проводитись класифікація
- 1.2. Додати ще один стовбець до матриці ознак зі значеннями 1
- 1.3. Встановити значення цільового вектору +1 та -1

$$y = \{-1, 1\}$$

Метод опорних векторів (SVM) в задачах класифікації

Розділення вибірки на навчальну та тестову

$$x_{train} = [0 \dots \ell]; \quad x_{test} = [\ell \dots N]$$

Побудова алгоритму навчання

Відступ $M_i(w) = y_i \langle x_i, w \rangle$

Задача оптимізації $f(w) = \frac{1}{\ell} \sum_{i=1}^{\ell} \left[\frac{1}{2} \|w\|^2 + C(1 - y_i \langle w, x_i \rangle)_+ \right]$

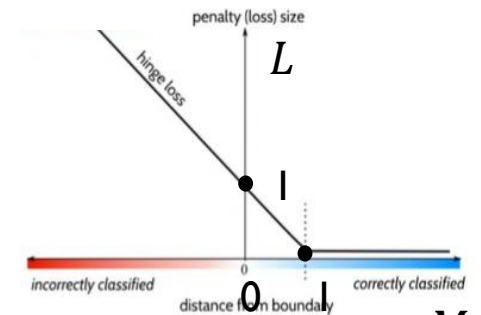
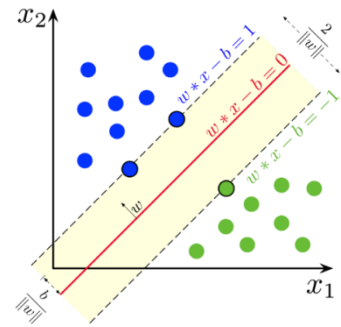
Розрахунок наступного значення ваг w за методом градієнтного спуску:

$$w_i = w_i - \eta \frac{1}{\ell} \nabla f(w)$$

Похідна від функціоналу якості

$$\nabla f(w) = \frac{1}{\ell} \sum_{i=1}^{\ell} \begin{cases} w, & \text{if } \max(1 - y_i \langle w, x_i \rangle)_+ = 0 \\ w - C y_i x_i, & \text{otherwise} \end{cases}$$

Визначити оптимальні значення параметра регуляризації C та швидкості навчання η для навчання на навчальній вибірці

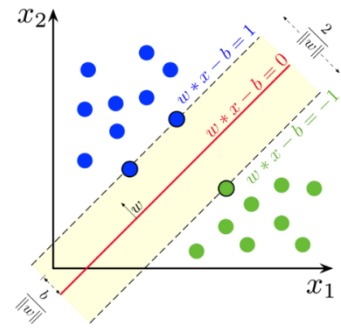


M_i

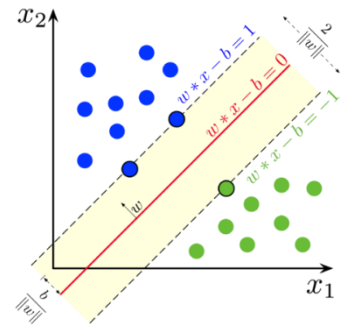
Метод опорних векторів (SVM)

Етапи розв'язання

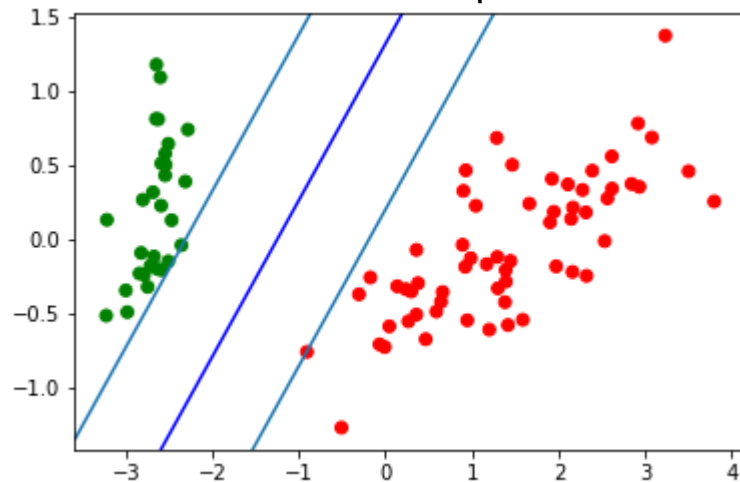
1. Імпортувати вибірку для проведення навчання
2. Розділити всю вибірку на навчальну та тестову
3. Провести підготовку даних до класифікації
 - визначити дві головних ознаки за якими буде проводитись класифікація додати ще один стовбчик до матриці ознак та встановити всі значення рівним 1
 - встановити значення цільового вектора +1 та -1
4. Побудувати алгоритм навчання на навчальній вибірці
 - встановити значення кроку навчання, коефіцієнта змінення ваг
 - визначити умову завершення навчання (встановити кількість епох навчання або мінімально допустиму точність алгоритму)
 - в залежності від величини зазору змінювати ваги за допомогою градієнта функції втрат Q
 - рахувати кількість помилок (неправильно класифікованих об'єктів) у процесі навчання
5. Подати графічно результат класифікації для навчальної вибірки та залежність помилок від номеру епохи навчання
6. Перевірити точність роботи алгоритму на тестовій вибірці
7. Подати графічно результат класифікації для тестової вибірки
8. Порівняти результати з SVM з sklearn
9. Оформити результати у вигляді звіту.



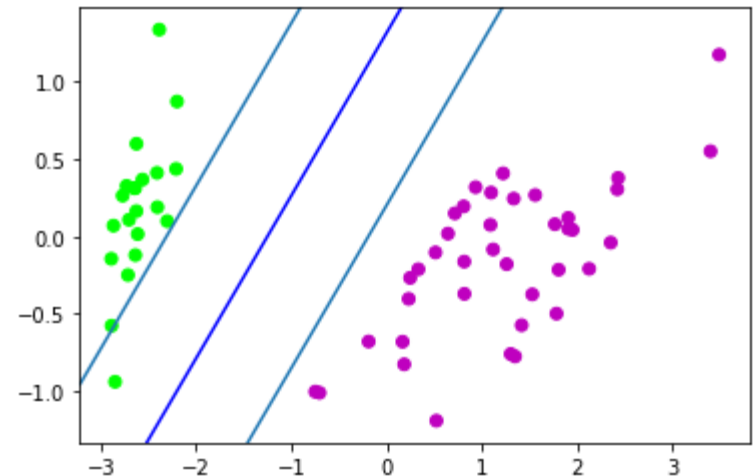
Приклад роботи програми на навчальній вибірці



Навчальна вибірка

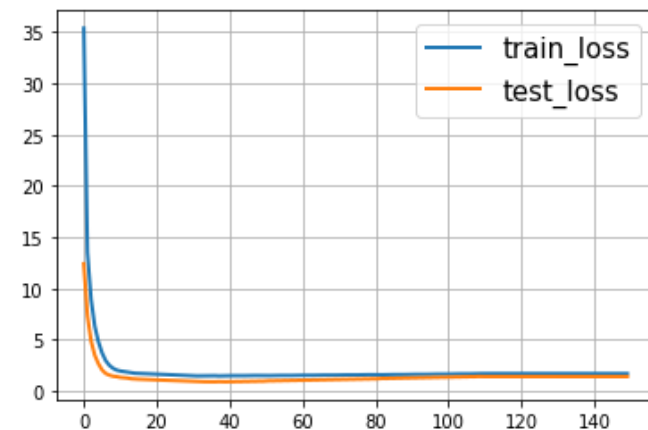


Тестова вибірка



Залежність помилки від епохи

```
[65 32 24 19 14 15 13 10 9 7 6 5 5 6 5 4 4 4 4 4 4 4 4  
4 4 4 4 4 4 4 4 3 3 3 3 3 2 2 2 1 2 1 2 1 2 1 2  
1 2 1 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1  
2 1 2 1 2 1 2 1 2 1 1 2 1 2 1 2 1 2 1 2 1 2 1 2  
1 2 1 2 1 2 1 2 1 2 1 2 1 2 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0]  
[ 0.96553668 -0.85848583 1.22764817]
```

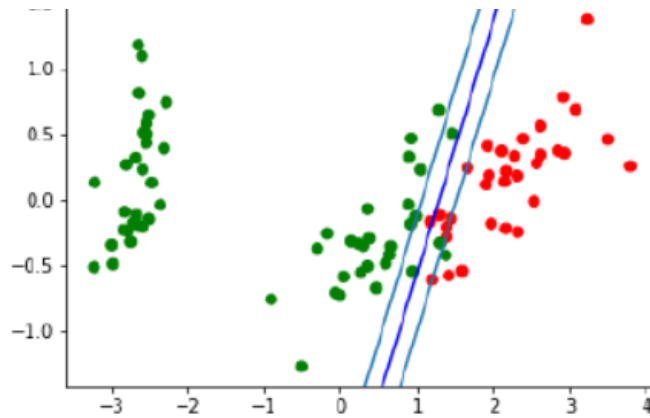
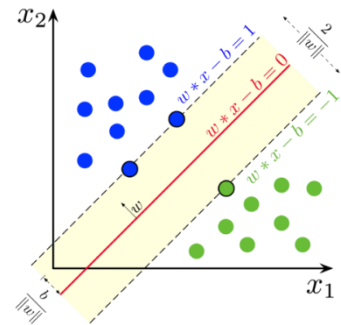


Метод опорних векторів (SVM)

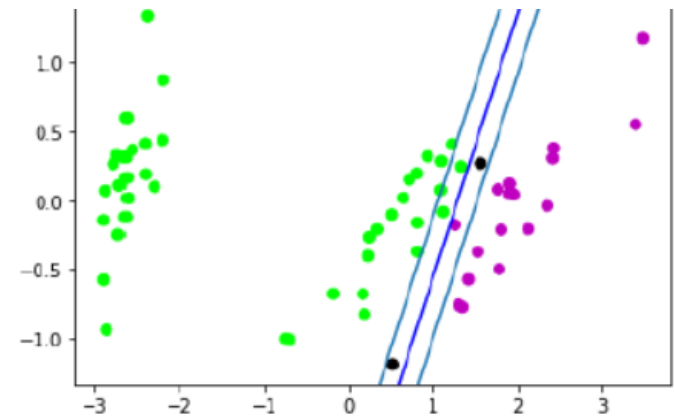
в задачах класифікації

Приклад роботи програми на навчальній вибірці

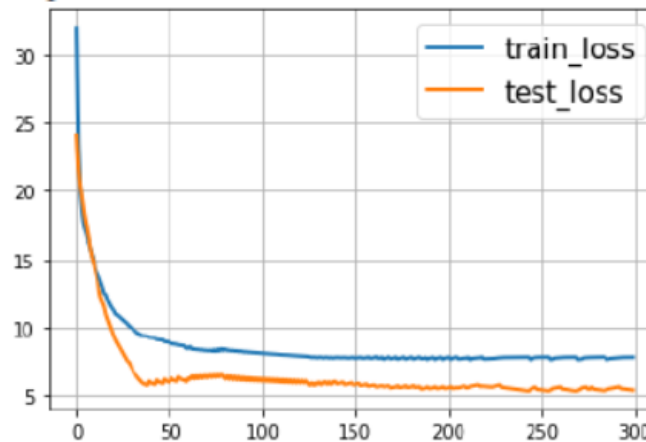
Результат класифікації для лінійно не роздільної вибірки



Навчальна вибірка



Тестова вибірка



Помилки класифікації

Метод опорних векторів (svm)

Постановка задачі на 60 балів

1. Розділити всю вибірку на навчальну та тестову
2. Побудувати алгоритм навчання на навчальній вибірці з використанням вбудованого алгоритму `SVC()` з бібліотеки `sklearn`
3. Перевірити результати роботи на тестовій вибірці.
4. Подати графічно результат класифікації для навчальної вибірки та залежність помилок від номеру епохи навчання
5. Оформити результати у вигляді звіту.