

ОСНОВИ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ, НЕЙРОННИХ МЕРЕЖ та ГЛИБОКОГО НАВЧАННЯ

Модуль 2. Навчання з вчителем

Лекція 2.8.

Класифікація. Метод SVM.

Класифікація

Формально:

Маємо множину \mathbb{O} об'єктів $o^{(j)}$, $j = 1, 2, \dots, M$

Кожен об'єкт $o^{(i)}$ має сукупність характеристик - ознак $x_i^{(j)}$, $i = 1, 2, \dots, N$ з множини \mathbb{X} .

Маємо множину \mathbb{C} класів $c^{(k)}$, $k = 2, \dots, K$

Існує невідома залежність (правило) \mathbb{F} , яка на підставі пар $\langle o^{(j)}, c^{(k)} \rangle$ визначає, чи належить об'єкт $o^{(j)}$ до класу $c^{(k)}$.

Завдання: знайти правило $\check{\mathbb{F}}$, максимально наближене до \mathbb{F} . Тобто, знайти вирішальне правило, що дозволяє класифікувати довільний об'єкт o за його ознаками.

Методи Класифікації

- Регресійні методи, логістична регресія
- Метод k-найближчих сусідів (KNN)
- *Метод опорних векторів (SVM)*
- Наївний Байєс (ймовірнісний класифікатор)
- Дерева рішень
-

Машина опорних векторів

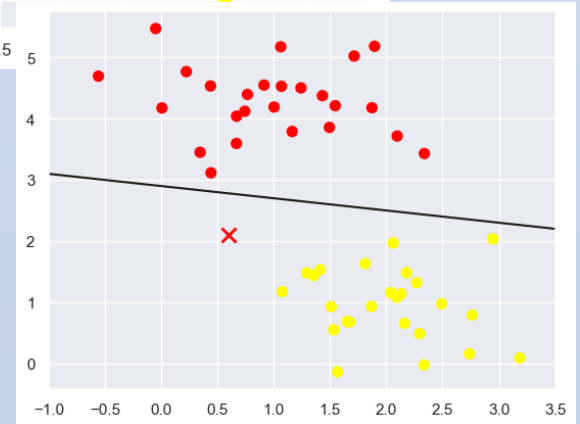
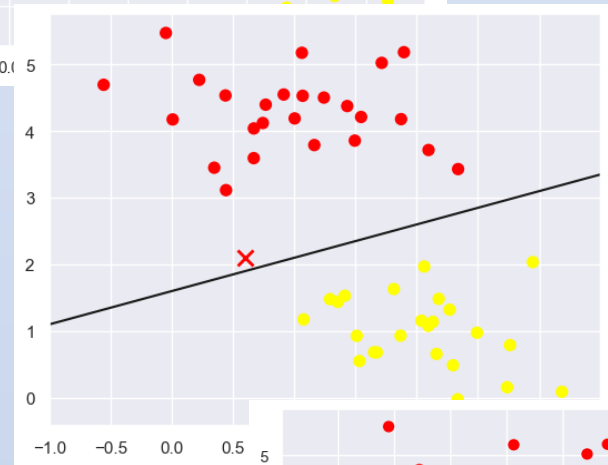
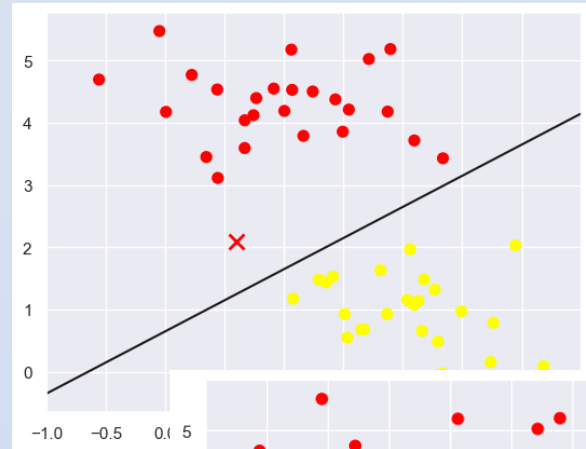
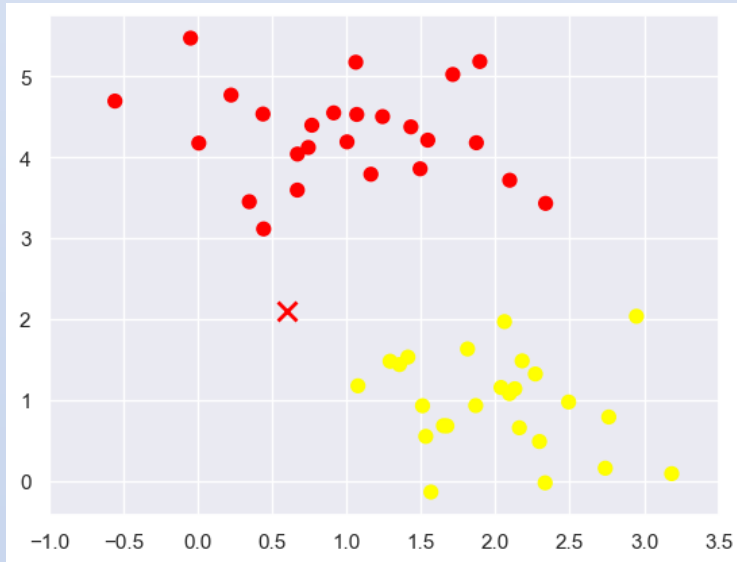
Метод опорних векторів (Support Vector machine, SVM, *опорно-векторна машина*) — комплекс алгоритмів навчання з учителем.

Основна ідея: Якщо розмірність простору ознак N , будується гіперплощина (розмірність $N-1$), що розділяє об'єкти вибірки деяким оптимальним (?) способом.

Особлива властивість SVM - безперервне зменшення емпіричної помилки класифікації та збільшення зазору, тому метод також відомий як метод класифікатора з максимальним зазором.

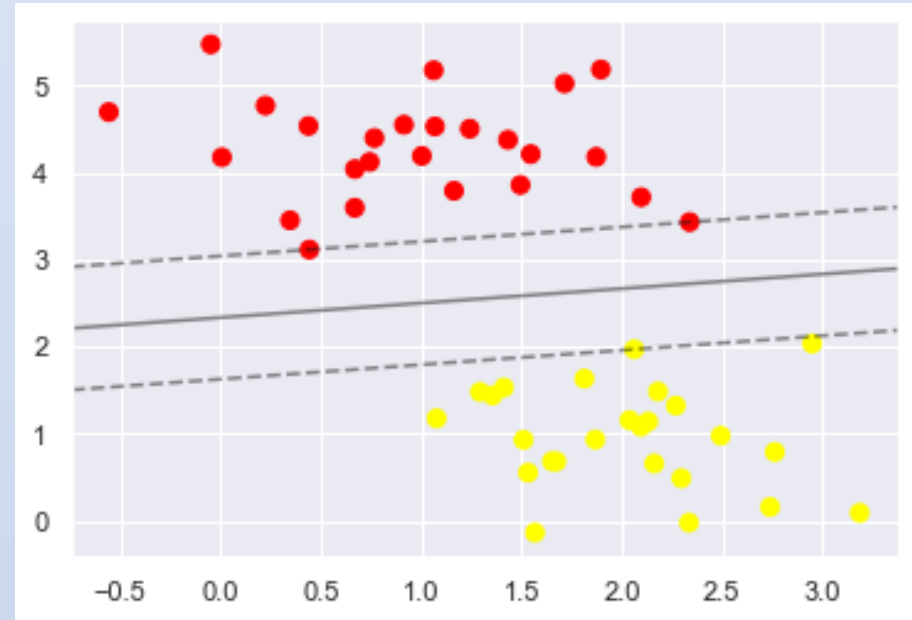
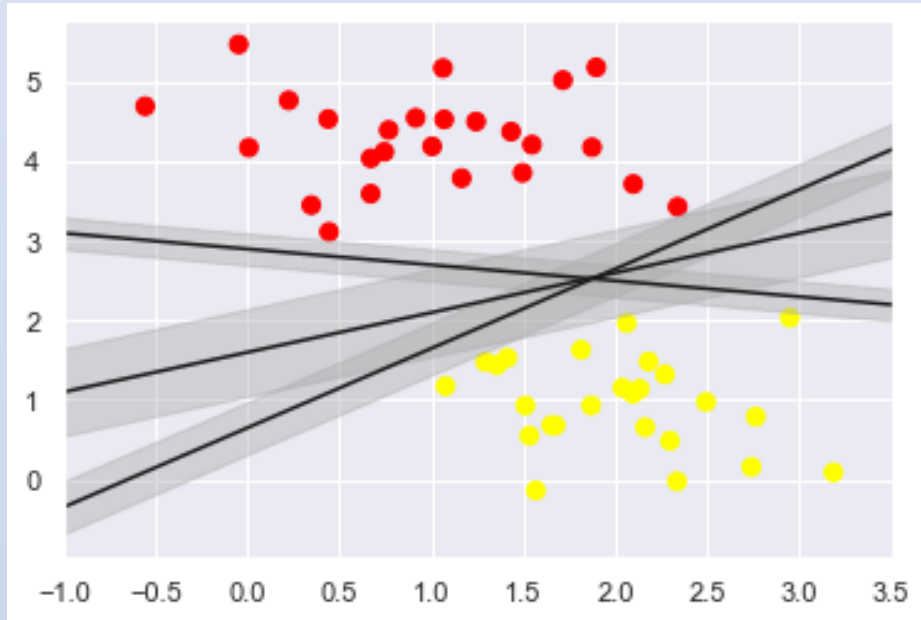
Машина опорних векторів

Приклад. Двовимірний простір, дві сепарабельні (лінійно-розділені) множини



Машина опорних векторів

Приклад. Двовимірний простір, дві сепарабельні (лінійно-розділені) множини



Обираємо пряму, що має найбільший зазор (проміжок)

Бінарна класифікація

Формально:

Маємо множину \mathcal{O} об'єктів $o^{(j)}$, $j = 1, 2, \dots, M$

Кожен об'єкт $o^{(i)}$ має сукупність характеристик - ознак $x_i^{(j)}$, $i = 1, 2, \dots, N$ з множини \mathbb{X} . Тобто $x^{(j)} \in \mathbb{R}^N$ - N -вимірний вектор.

Маємо два класи $c^{(0)} \rightarrow y = -1$, $c^{(1)} \rightarrow y = 1$.

Завдання: знайти правило f , що за довільним вектором $x \in \mathbb{R}^N$ дозволяє визначити його клас, тобто повертає -1 , або $+1$.

Бінарна класифікація

Два лінійно сепарабельні класи.

Правило F шукається як функція

$$F(x) = \text{sign}(\langle W, x \rangle + b)$$

де W – нормальний вектор до гіперплощини, b – зсув гіперплощини від початку координат.

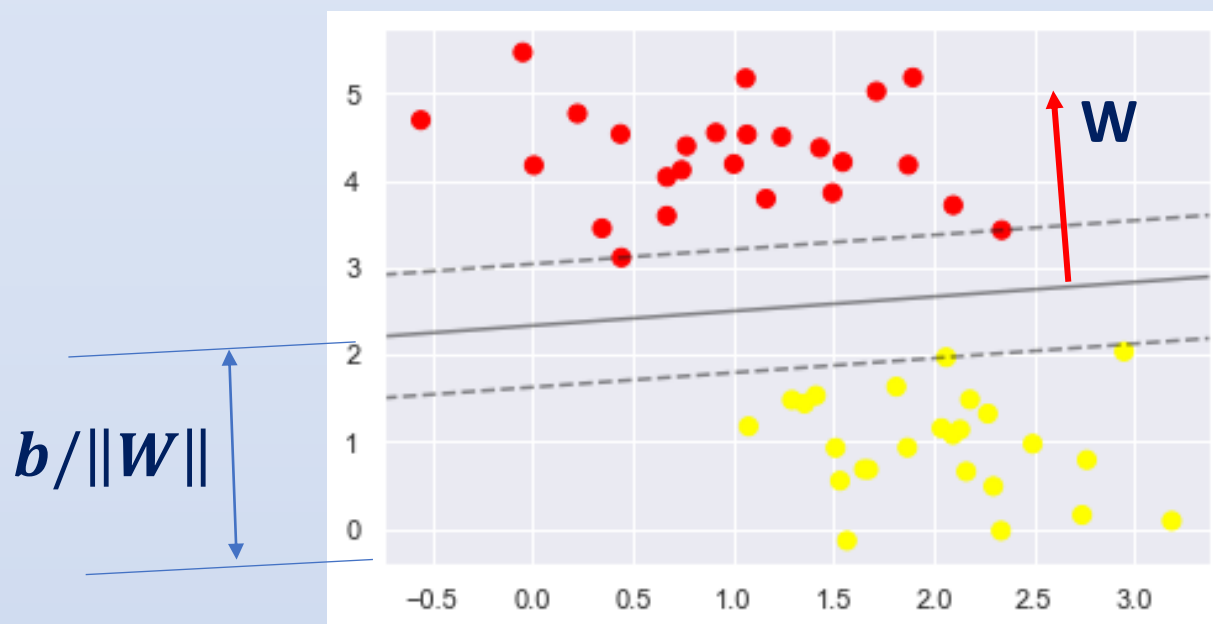
Обираються дві паралельні гіперплощини, які розділяють два класи даних так, що відстань між ними якомога більша. Ці дві гіперплощини може бути описано рівняннями:

$$\langle W, x \rangle + b = 1$$

$$\langle W, x \rangle + b = -1$$

Область, обмежена цими двома гіперплощинами, називається «розділенням» (margin), а максимально розділова гіперплощина лежить посередині між цими двома.

Бінарна класифікація



$$Wx + b = 1$$

$$Wx + b = 0$$

$$Wx + b = -1$$

Відстань між гіперплощинами: $margin = 2/\|W\|$

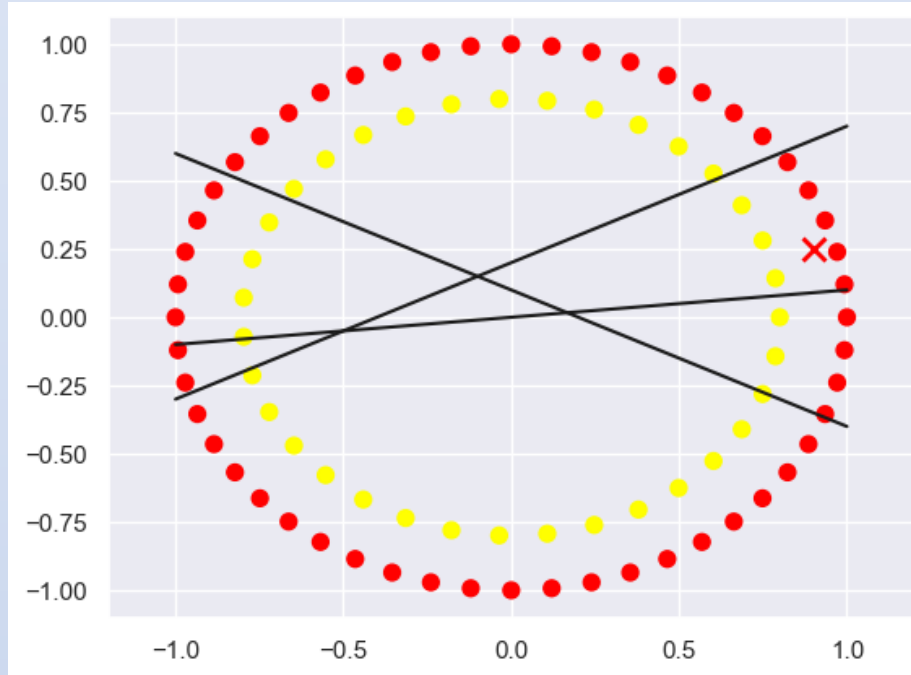
Тобто необхідно мінімізувати $\|W\|$ при виконанні обмежень

$$Wx^{(j)} + b \geq 1, \quad \forall x^{(j)} \in c^{(1)}$$

$$Wx^{(j)} + b \leq -1, \quad \forall x^{(j)} \in c^{(0)}$$

Машина опорних векторів

Лінійно НЕ сепарабельні класи.



Загальна ідея: все об'єкти вкладаються у простір \mathbb{R}^{N+1} – простір вищої розмірності з допомогою спеціального відображення $\varphi: \mathbb{R}^N \rightarrow \mathbb{R}^{N+1}$.

При цьому відображення φ вибирається так, щоб у новому просторі об'єкти були розділені лінійно.

Машина опорних векторів

Лінійно НЕ сепарабельні класи.

Функцію, що розділяє класи шукають у вигляді

$$F(x) = \text{sign} (\langle W, \varphi(x) \rangle + b)$$

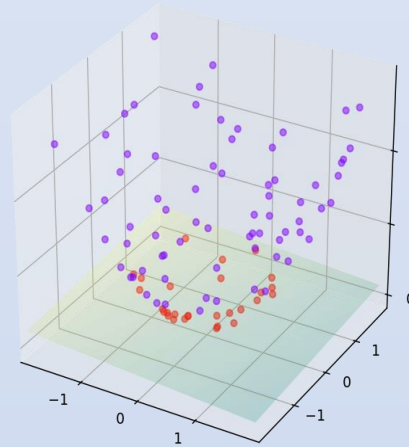
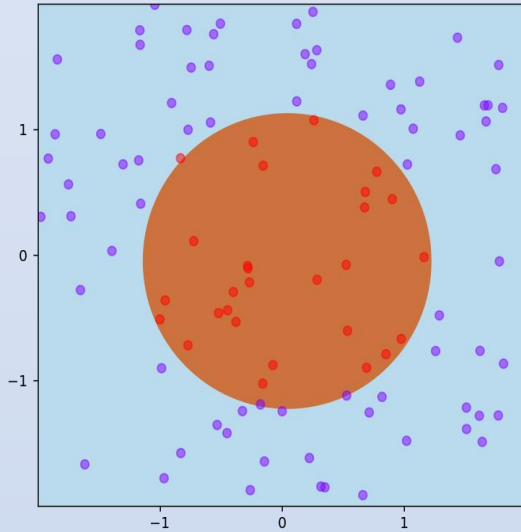
Функція $\varphi(\cdot)$ - ядро (kernel) класифікатора.
Як обрати $\varphi(\cdot)$???

Типові ядра:

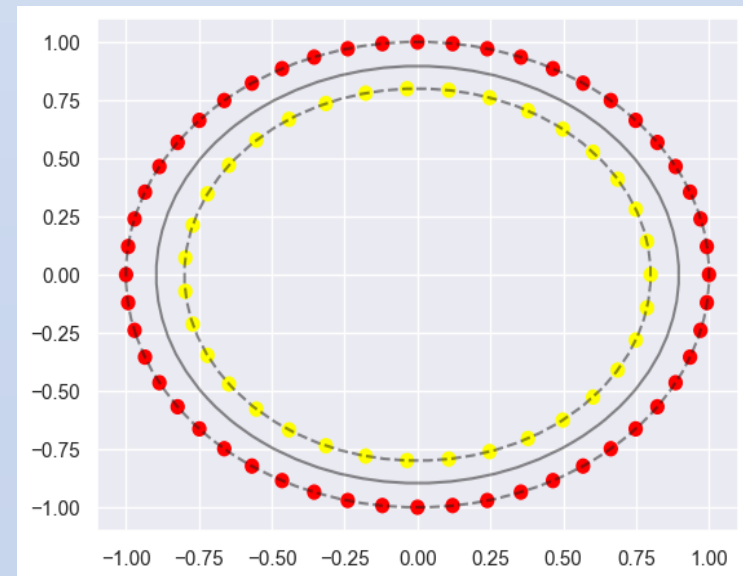
- Лінійне ('linear')
- Поліноміальне ('poly') $\varphi(x, x') = (\langle x, x' \rangle + \text{const})^d, d > 1$
- Сігмоїд ('sigmoid') $\varphi(x, x') = \tanh(k\langle x, x' \rangle + c)$, $k > 0, c > 0$
- Експоненційне ('rbf') ...
- ...

Машина опорних векторів

Поліноміальне ядро ($d=2$)



Гіперплощини для
`kernel="poly", degree=2`



Бінарна класифікація

Scikit-learn: модуль `Scikit-learn.svm` включає наступні SVM алгоритми

<code>svm.LinearSVC([penalty, loss, dual, tol, C, ...])</code>	Linear Support Vector Classification.
<code>svm.LinearSVR(*[, epsilon, tol, C, loss, ...])</code>	Linear Support Vector Regression.
<code>svm.NuSVC(*[, nu, kernel, degree, gamma, ...])</code>	Nu-Support Vector Classification.
<code>svm.NuSVR(*[, nu, C, kernel, degree, gamma, ...])</code>	Nu Support Vector Regression.
<code>svm.OneClassSVM(*[, kernel, degree, gamma, ...])</code>	Unsupervised Outlier Detection.
<code>svm.SVC(*[, C, kernel, degree, gamma, ...])</code>	C-Support Vector Classification.
<code>svm.SVR(*[, kernel, degree, gamma, coef0, ...])</code>	Epsilon-Support Vector Regression.

Приклади дивись:

Приклад 1. `lec_02_08_Exmpl_1.md`

Приклад 2. `lec_02_08_Exmpl_2.md`

Переваги та недоліки SVM

Переваги SVM перед методом стохастичного градієнта та нейронними мережами:

- Завдання опуклого квадратичного програмування добре вивчене і має єдине рішення.
- Метод опорних векторів еквівалентний двошаровій нейронній мережі, де число нейронів на прихованому шарі визначається автоматично як число опорних векторів.
- Принцип оптимальної роздільної гіперплощини призводить до максимізації ширини смуги, що розділяє, а отже, до більш впевненої класифікації.

Переваги та недоліки SVM

Недоліки класичного SVM:

- **Нестійкість до шуму:** викиди у вихідних даних стають опорними об'єктами-порушниками та безпосередньо впливають на побудову роздільної гіперплощини.
- **Не описані загальні методи побудови ядер та спрямовуючих просторів,** що найбільш підходять для конкретного завдання.
- **Немає відбору ознак.**
- **Необхідно підбирати константу C за допомогою крос-валідації.**

Контрольні запитання

- **Надайте загальну постановку задачі класифікації.**
- **Пояснить сутність методу SVM вирішення задачі класифікації.**
- **Пояснить, як вирішується завдання класифікації коли класи не сепарабельні.**

Корисні та цікави посилання

- **Машинне навчання**

https://uk.wikipedia.org/wiki/машинне_навчання

- **Львівська політехніка**

<http://www.mmf.lnu.edu.ua/ar/1739>

<http://www.mmf.lnu.edu.ua/ar/1743>

Рекомендована ЛІТЕРАТУРА

- **Глибинне навчання:** Навчальний посібник / Уклад.: В.В. Литвин, Р.М. Пелещак, В.А. Висоцька В.А. – Львів: Видавництво Львівської політехніки, 2021. – 264 с.
- Тимощук П. В., Лобур М. В. **Principles of Artificial Neural Networks and Their Applications: Принципи штучних нейронних мереж та їх застосування:** Навчальний посібник. – Львів : Видавництво Львівської політехніки, 2020. – 292 с.
- Morales M. **Grokking Deep Reinforcement Learning.** – Manning, 2020. – 907 с.
- Trask Andrew W. **Grokking Deep Learning.** – Manning, 2019. – 336 с.

The END

Модуль 2. Лекція 2.8.