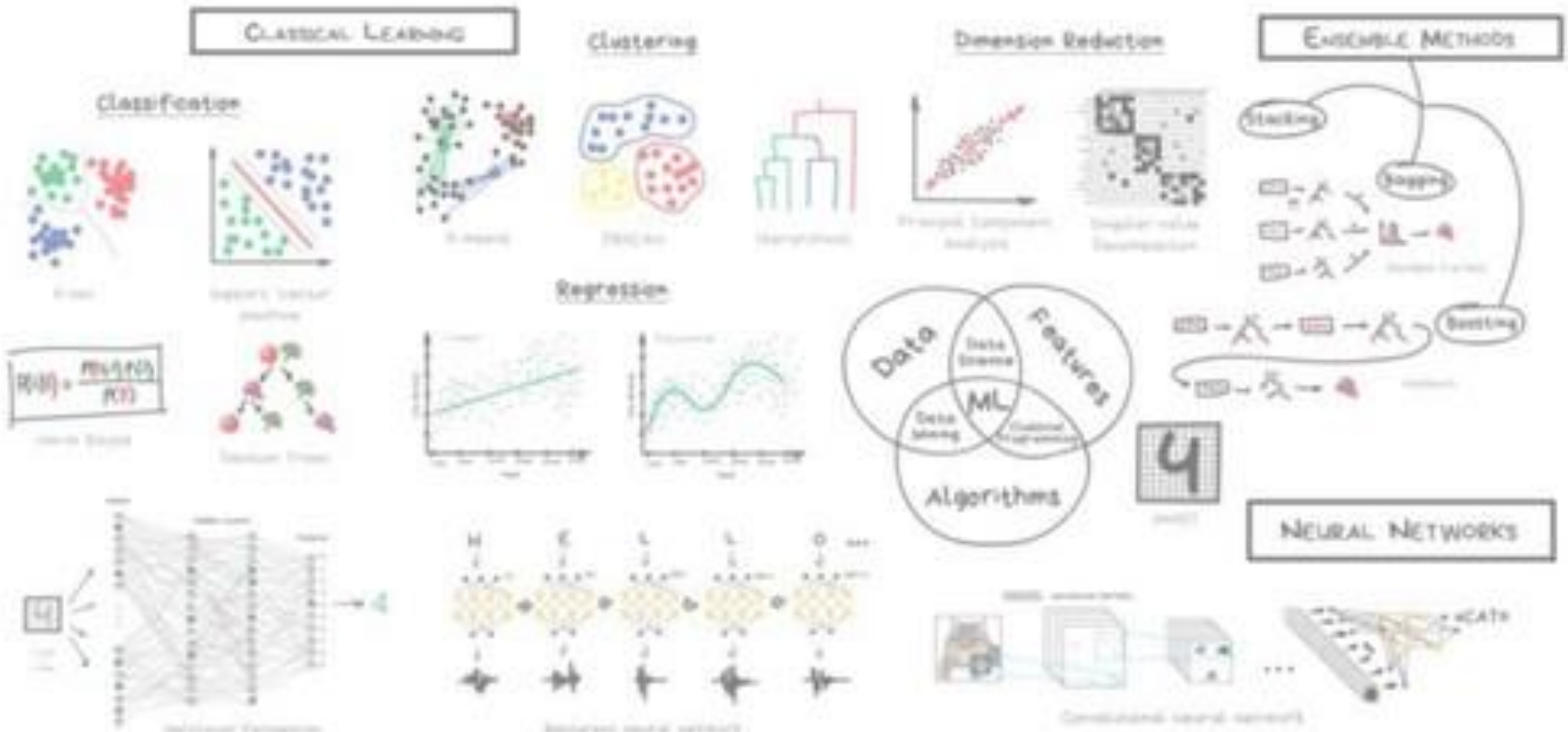
МАШИНЕ НАВЧАННЯ

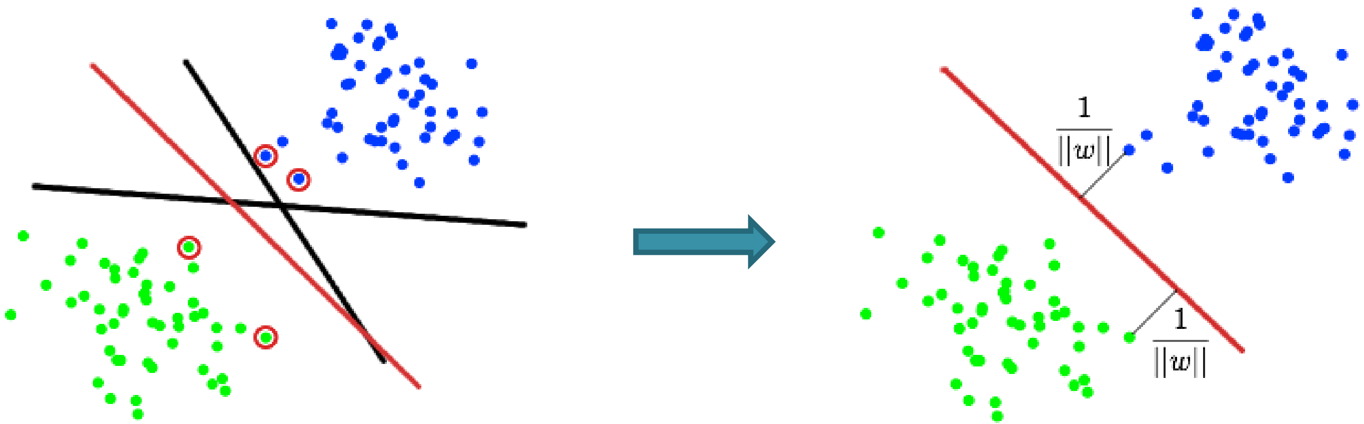
Класичне навчання. Навчання з вчителем



Лекція №6

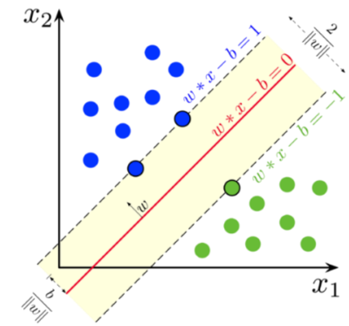
Методи класифікації
Метод опорних векторів (SVM)

Найбільш популярний метод класичної класифікації. Ним класифікували вже **все**: типи рослин, лиця на фотографіях, документи за тематиками. Багато років він був головною відповіддю на питання «який би мені взяти класифікатор».



Ідея SVM — провести полосу пряму між категоріями таким чином, що відстані від неї до граничного об’єкта кожного класу була максимальною.

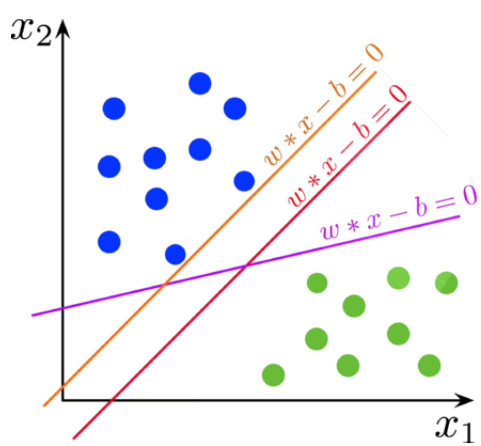
Метод опорних векторів (SVM) в задачах класифікації



Розглянемо задачу бінарної класифікації, в якій об'єктам з $X = \mathbb{R}^n$ (об'єкти описуються n числовими признаками) відповідає один з двох класів $Y = \{-1, +1\}$.
Нехай задана навчальна вибірка пар "об'єкт-відповідь": $(x_i, yi), i = 1 \dots \ell$.
Необхідно побудувати алгоритм класифікації $a(x): X \to Y$.

Розділяюча гіперплощина

У просторі $\mathbb{R}^n$ рівняння $\langle w, x \rangle - b = 0$ при заданих w та b визначає гіперплощину, що розділяє $\mathbb{R}^n$ на два класи: C_1 та C_2:



$$\begin{cases} \langle w, x \rangle - b > 0, & \forall x \in C_1 \\ \langle w, x \rangle - b < 0, & \forall x \in C_2 \end{cases}$$

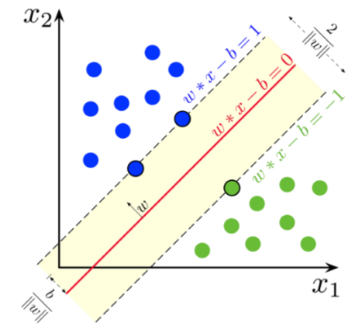
або

$$\begin{cases} \langle w, x \rangle - b < 0, & \forall x \in C_1 \\ \langle w, x \rangle - b > 0, & \forall x \in C_2 \end{cases}$$

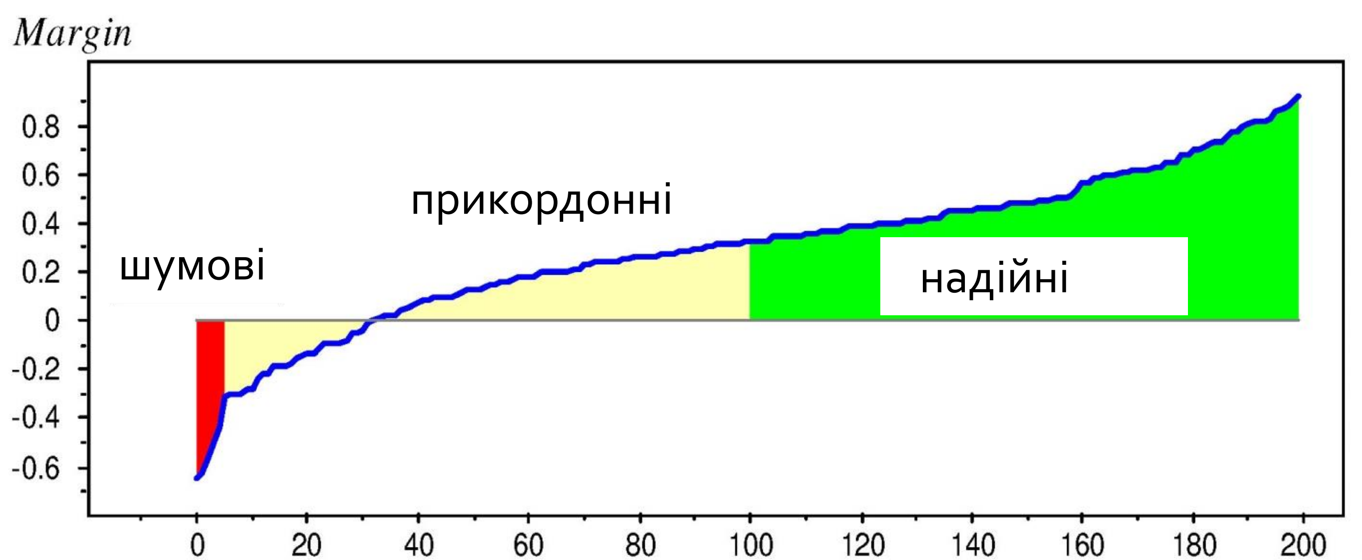
w — вектор нормалі до гіперплощини

$\frac{b}{\|w\|}$ — відстань від гіперплощини до початку координат

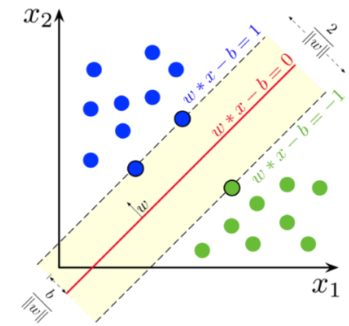
Метод опорних векторів (SVM) в задачах класифікації:
Лінійно роздільна вибірка



Відступ (Margin) — характеристика, яка оцінює, наскільки об’єкт «занурений у свій клас, наскільки типовим представником свого класу він є.
Чим менше значення відступа M_i, тим ближче об’єкт x_i підходе до границі класів і тим вище стає ймовірність помилки.
Відступ M_i від’ємний лише тоді, коли алгоритм класифікації $a(x_i)$ допускає помилку на об’єкті x_i .



Метод опорних векторів (SVM) в задачах класифікації:
Лінійно роздільна вибірка



Для **лінійного класифікатора** відступ визначається рівнянням:

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

Для **лінійно роздільної вибірки** існує така гіперплощина, відступ від якої до кожного об'єкта є позитивним:

$$\exists w, b: \quad Mi(w, b) = yi(\langle x_i, w\rangle - b) > 0, \qquad i = 1 \dots \ell$$

Задача: поюудувати таку розділяючу гіперплощину, щоб об'єкти навчальної вибірки знаходились на найбільшій відстані від неї.

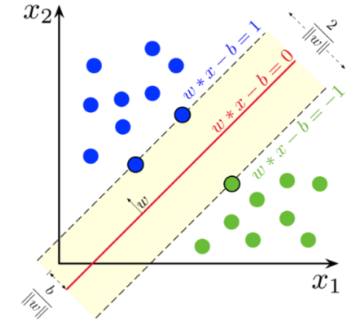
Нормування

При множенні w та b на константу $C \neq 0$ рівняння $\langle x_i, Cw\rangle - Cb = 0$ визначає ту ж саму гіперплощину, що й $\langle x_i, w\rangle - b = 0$.

Для зручності проведемо нормування: оберемо константу C таким чином, щоб $\min M_i\,(w, b) = 1$.

Метод опорних векторів (SVM) в задачах класифікації:

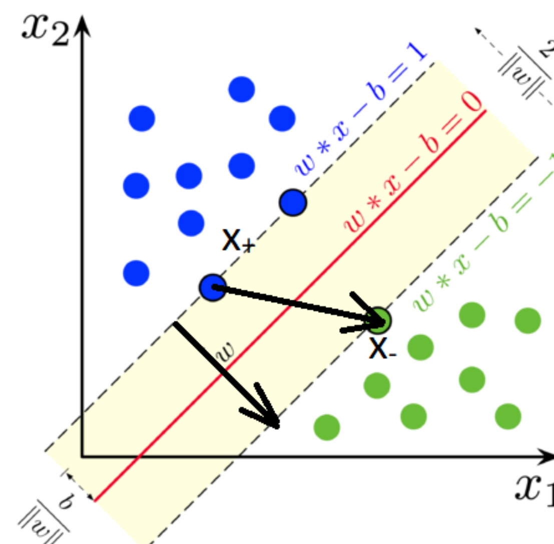
Лінійно роздільна вибірка



В кожному з двох класів нзайдеться хоча б один "граничний" об’єкт навчальної вибірки, відступ якого дорівнює цьому мінімуму: в іншому випадку можна було б змістити гіперплощину в бік класу з більшим відступом, тим самим збільшити мінімальну відстанб від гіперплощини до об’єктів навчальної вибірки.

Позначимо "граничний" об’єкт з класа $+1$ як x_+ , а з класа -1 як x_- .

$$x_+\colon \langle x, w\rangle - b = +1$$
$$x_-\colon \langle x, w\rangle - b = -1$$
$$M_+(w,b) = (+1)(\langle x_+, w\rangle - b) = 1$$
$$M_-(w,b) = (-1)(\langle x_-, w\rangle - b) = 1$$



Нормування дозволяє обмежити розділяючу полосу між класами:

$$x\colon \{-1 < \langle x, w\rangle - b < 1\}$$

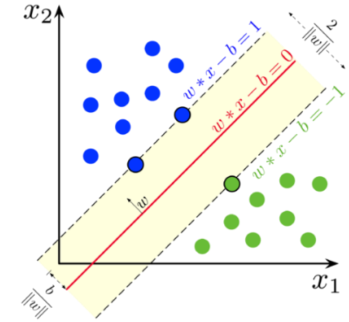
Всередині якої не може лежати жоден об’єкт Навчальної вибікри.

Ширина полоси це проекція вектора $(x_+ - x_-)$ на w:

$$\frac{\langle x_+ - x_-, w\rangle}{\|w\|} = \frac{M_+ + M_-}{\|w\|} = \frac{2}{\|w\|} \to \max \Longrightarrow \|w\| \longrightarrow \min.$$

$\|w\| = \langle w, w\rangle$ – скалярний добуток

Метод опорних векторів (SVM) в задачах класифікації:
Лінійно роздільна вибірка



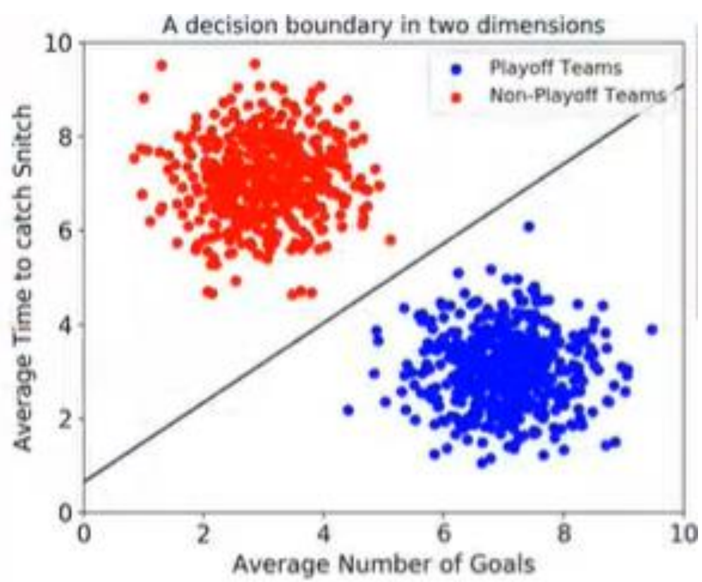
Максимальність ширини розділяючої полоси

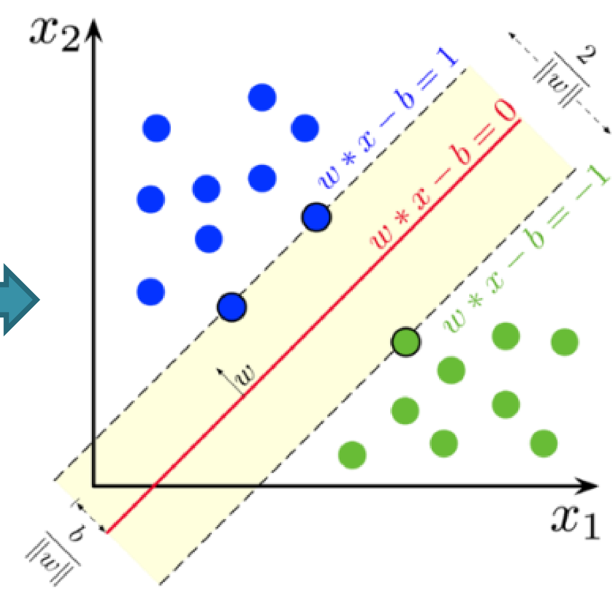
$$\frac{2}{\|w\|} \to \max \Longrightarrow \|w\| \longrightarrow \min.$$

$\|w\| = \langle w, w \rangle$ – скалярний добуток

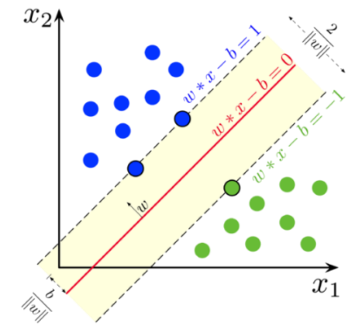
Постановка задачи оптимизации в терминах квадратичного программирования:

$$\begin{cases} \|w\| \quad \to \min\limits_{w,b} \\ M_i(w,b) \geq 1, \qquad i = 1 \dots \ell \end{cases}$$





Метод опорних векторів (SVM) в задачах класифікації



Умови Каруша-Кунна-Таккера

задача нелінійного програмування з обмеженнями:

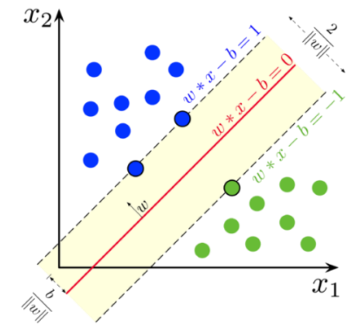
$$\begin{cases} f(x) \longrightarrow \min\limits_{x \in X} \\ g_i(x) \leq 0, \qquad i = 1 \dots m \\ h_j(x) = 0, \qquad j = 1 \dots k \end{cases}$$

Якщо x — точка локального мінімума при накладених обмеженнях, то існують такі множники $\mu_i, i = 1 \dots m$ та $\lambda_j, j = 1 \dots k$, що для функції Лагранжа $\mathcal{L}(x; \mu, \lambda)$ виконуються умови:

$$\begin{cases} \dfrac{d\mathcal{L}}{dx} = 0, \mathcal{L}(x; \mu, \lambda) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{k} \lambda_j h_j(x) \\ g_i(x) \leq 0, \qquad h_j(x) = 0 \qquad \text{(вихідні обмеження)} \\ \mu_i \geq 0 \qquad \text{(двійкові обмеження)} \\ \mu_i g_i(x) = 0 \qquad \text{(доповнюючі обмеження)} \end{cases}$$

При цьому шукана точка є сідловою точкою функції Лагранжа: мінімумом по x та максимумом по двійковим змінним μ.

Метод опорних векторів (SVM) в задачах класифікації



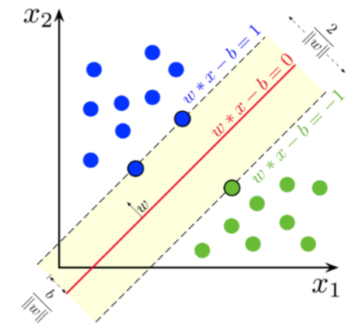
Для нашої задачі оптимізації:
$$\begin{cases} \|w\| \quad \to \min\limits_{w,b} \\ M_i(w,b) \geq 1, \quad i = 1 \dots \ell \end{cases}$$

За теоремою ККТ маємо Лагранжиан
$$\mathcal{L}(w, b, \lambda_1, \dots, \lambda_\ell) = \frac{1}{2}\|w\| - \sum_{i=1}^{\ell} \lambda_i \{y_i(\langle x_i, w\rangle - b) - 1\}$$

Похідні від Лагранжиану за параметрами
$$\begin{cases} \dfrac{\partial \mathcal{L}(w, b, \lambda_1, \dots, \lambda_\ell)}{\partial w} = 0; \\ \dfrac{\partial \mathcal{L}(w, b, \lambda_1, \dots, \lambda_\ell)}{\partial b} = 0; \\ \lambda_i \geq 0; \\ \lambda_i = 0 \text{ або } \{y_i(\langle x_i, w\rangle - b) - 1\} = 0 \end{cases}$$

З перших двох рівнянь знаходимо:
$$\begin{cases} w = \sum_{i=1}^{\ell} \lambda_i y_i x_i; \\ \sum_{i=1}^{\ell} \lambda_i y_i = 0 \end{cases}$$

Метод опорних векторів (SVM) в задачах класифікації



Підставляючи отримані обмеження в функцію Лагранжа отримуємо постановку двійкової задачі, яка залежить лише від двійкових змінних λ:

$$\begin{cases} \mathcal{L}(\lambda) = \sum_{i=1}^{\ell} \lambda_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle \\ 0 \le \lambda_i \le C, \qquad i = 1 \dots \ell \\ \sum_{i=1}^{\ell} \lambda_i y_i = 0 \end{cases}$$

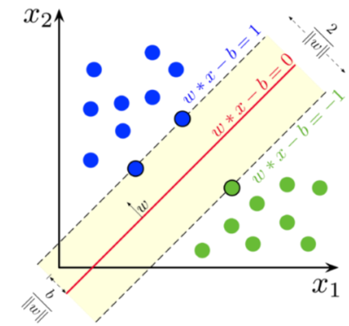
$$\mathcal{L}(\lambda) = \sum_{i=1}^{\ell} \lambda_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle =$$

$$\{(H)_{ij} = y_i y_j \langle x_i, x_j \rangle\}$$

$$= \sum_{i=1}^{\ell} \lambda_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \lambda_i \lambda_j \, (H)_{ij} = \boldsymbol{\lambda} + \frac{1}{2} \boldsymbol{\lambda}^T \boldsymbol{H} \boldsymbol{\lambda}$$

Це задача квадратичного програмування:

Метод опорних векторів (SVM) в задачах класифікації



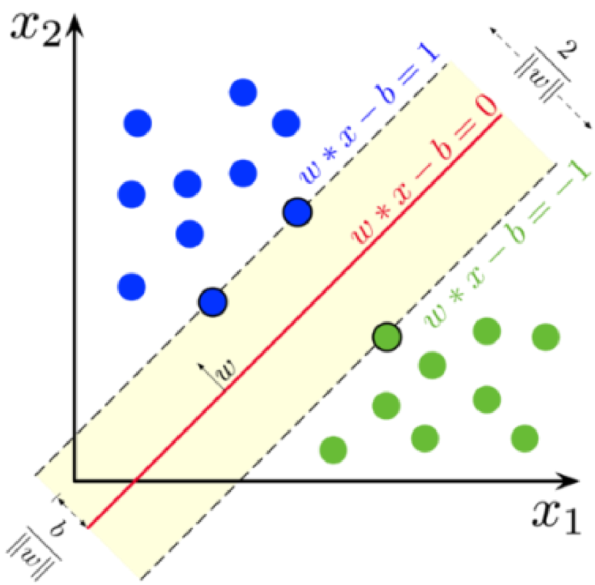
Метод градієнтного спуску:

$$\mathcal{L}(\lambda) = \boldsymbol{\lambda} + \frac{1}{2}\boldsymbol{\lambda}^T \boldsymbol{H}\boldsymbol{\lambda}$$

Невідомі параметри Лагранжа:

$$\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t + \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}}$$

Навчальна вибірка



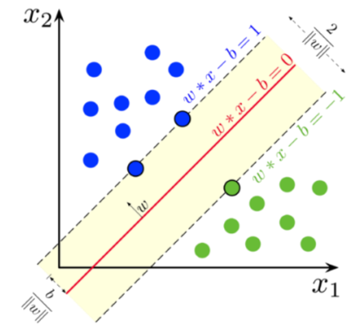
За отриманими λ_i розв'язок прямої задачі:

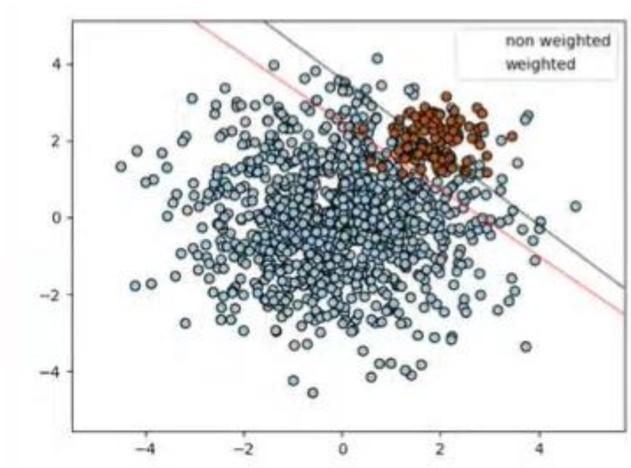
$$\begin{cases} w = \sum_{i=1}^{\ell} \lambda_i y_i x_i; \\ b = \begin{cases} \overline{(\langle w, x_i \rangle - y_i)} \\ med(\langle w, x_i \rangle - y_i) \end{cases} \forall i\colon \lambda_i > 0,\ M_i = 1 \end{cases}$$

Лінійний класифікатор

$$a(x) = sign\left(\sum_{i=1}^{\ell} \lambda_i y_i \langle x_i, x \rangle - b\right)$$

Метод опорних векторів (SVM) в задачах класифікації:
Лінійно не роздільна вибірка





В даних можуть бути певні викиди, що приводить до нечітких меж між класами. Необхідно послабити обмеження, дозволяючи де-яким об'єктам попадати в середину розділяючої полоси та на "територію" іншого класу.

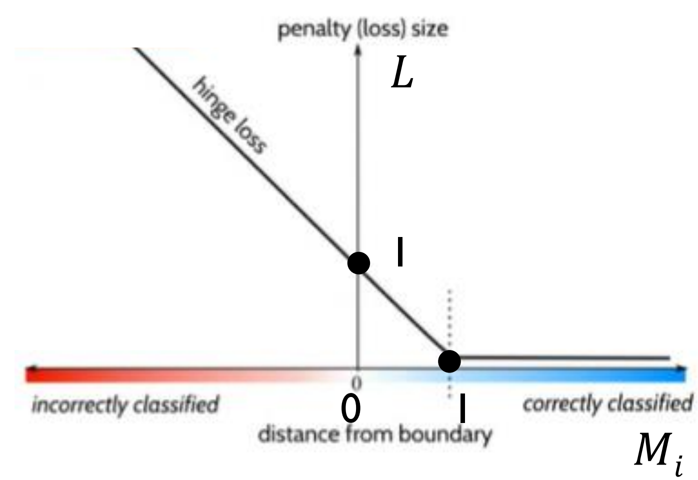
Відступ

$$M_i(w, b) = yi(\langle x_i, w\rangle - b)$$

Функція втрат (Loss function)

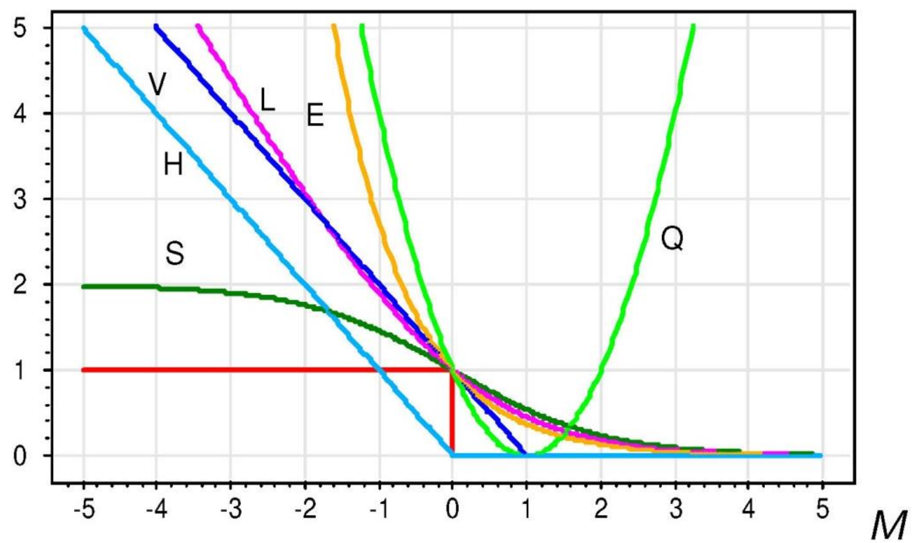
$$L = \max(0{,}1 - Mi)$$

Інакше всі помилки рівноправні



Неперервні апроксимації порогової функції втрат

Часто використовувані неперервні функції втрат $L(M)$



$V(M) = (1 - M)_+$ – кусочно-лінійна (SVM);

$H(M) = (-M)_+$ – кусочно-лінійна (Hebb's rule);

$L(M) = \log_2(1 + e^{-M})$ – логарифмічна (LR);

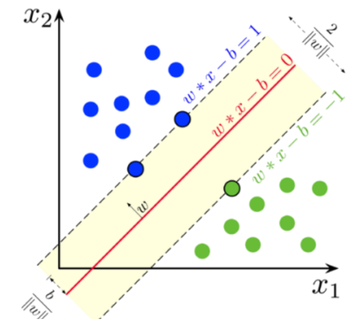
$Q(M) = (1 - M)^2$ – квадратична (FLD);

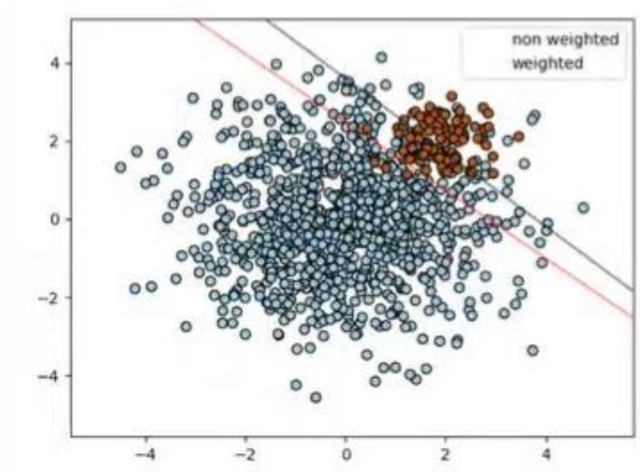
$S(M) = 2(1 + e^{M})^{-1}$ – сигмоїдна (ANN);

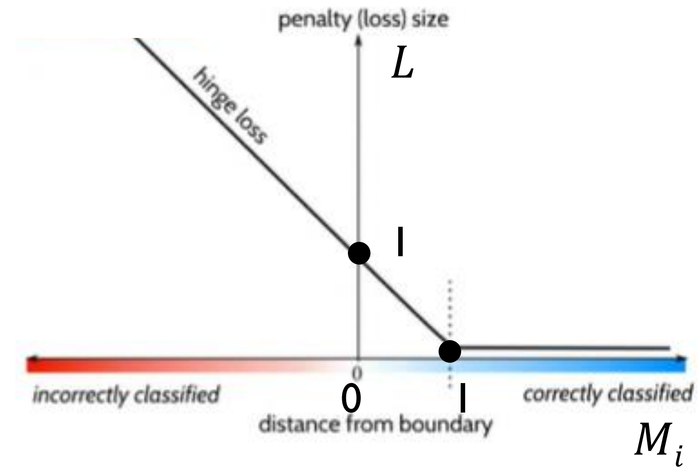
$E(M) = e^{-M}$ – експоненціальна (AdaBoost);

$[M < 0]$ – порогова функція втрат

Метод опорних векторів (SVM) в задачах класифікації:
Лінійно не роздільна вибірка







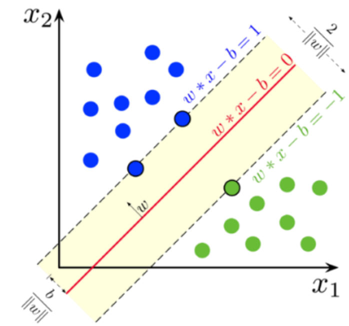
- Для кожного об'єкта віднімемо від відступа позитивну величину ξ_i, яле будемо вимагати, що ці уведені поправки були мінімальними. Це приведе до задачі, що має назву *SVM з м'яким відступом* (англ. *soft-margin SVM*):

$$\begin{cases} \frac{1}{2}\|w\| + C\sum_{i=1}^{\ell}\xi_i \longrightarrow \min\limits_{w,b,\xi} \\ M_i(w,b) \geq 1 - \xi_i, \qquad i = 1 \dots \ell \\ \xi_i \geq 0, \qquad i = 1 \dots \ell \end{cases}$$

- Еквівалентна задача безумовної мінімізації:

$$\frac{1}{2}\|w\| + C\sum_{i=1}^{\ell}(1 - M_i(w,b))_+ \longrightarrow \min_{w,b}$$

Метод опорних векторів (SVM) в задачах класифікації



За теоремою Каруша—Куна—Таккера, поставлена задача мінімізації еквівалентна двійковій задачі пошуку сідлової точки функції Лагранжа:

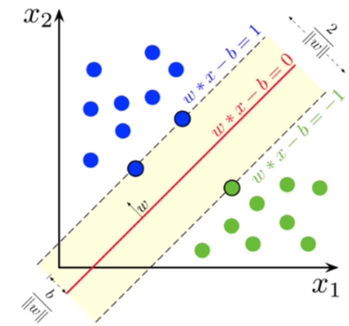
$$\mathcal{L}(w, b, \xi; \lambda, \eta) = \frac{1}{2}\|w\| - \sum_{i=1}^{\ell} \lambda_i (M_i(w, b) - 1) - \sum_{i=1}^{\ell} \xi_i \, (\lambda_i + \eta_i - C)$$

λ_i — змінні, двійкові до обмежень $M_i \geq 1 - \xi_i$
η_i — змінні, двійкові до обмежень $\xi_i \geq 0$

Необхідні умови існування сідлової точки функції Лагранжа:

$$\begin{cases} \dfrac{d\mathcal{L}}{dw} = 0, \quad \dfrac{d\mathcal{L}}{db} = 0, \quad \dfrac{d\mathcal{L}}{d\xi} = 0 \\ \xi_i \geq 0, \quad \lambda_i \geq 0, \quad \eta_i \geq 0, \qquad i = 1 \ldots \ell \\ \lambda_i M_i(w, b) = 0 \implies \lambda_i = 0 \text{ або } M_i(w, b) = 1 - \xi_i, \qquad i = 1 \ldots \ell \\ \eta_i \xi_i = 0 \implies \eta_i = 0 \text{ або } \xi_i = 0, \qquad i = 1 \ldots \ell \end{cases}$$

Метод опорних векторів (SVM) в задачах класифікації



Відступ (margin) об'єкта x_i від розділяючої гіперплощини :

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

Функція Лагранжа:

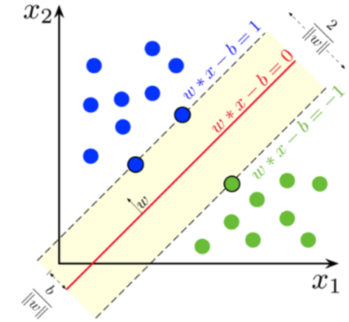
$$\mathcal{L}(w,b,\xi;\lambda,\eta) = \frac{1}{2}\|w\| - \sum_{i=1}^{\ell}\lambda_i(M_i(w,b)-1) - \sum_{i=1}^{\ell}\xi_i\,(\lambda_i+\eta_i - C)$$

Необхідні умови існування сідлової точки фугкції Лагранжа:

$$\begin{cases} \dfrac{d\mathcal{L}}{dw} = w - \sum_{i=1}^{\ell}\lambda_i y_i x_i = 0 \quad\Longrightarrow\quad w = \sum_{i=1}^{\ell}\lambda_i y_i x_i \\ \dfrac{d\mathcal{L}}{db} = \sum_{i=1}^{\ell}\lambda_i y_i = 0 \quad\Longrightarrow\quad \sum_{i=1}^{\ell}\lambda_i y_i = 0 \\ \dfrac{d\mathcal{L}}{d\xi_i} = -\lambda_i - \eta_i + C = 0 \quad\Longrightarrow\quad \eta_i + \lambda_i = C, \qquad i = 1\ldots\ell \end{cases}$$

Зауважимо, що $\eta_i \geq 0,\ \lambda_i \geq 0,\ C > 0$, тому, з останнього обмеження отримуємо $0 \leq \eta_i \leq C,\ \ 0 \leq \lambda_i \leq C$.

Метод опорних векторів (SVM) в задачах класифікації

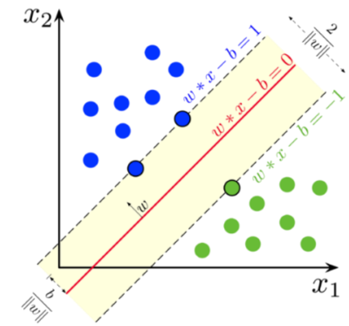


Система умов Каруша-Куна-Таккера:

$$\begin{cases} w = \sum_{i=1}^{\ell} \lambda_i y_i x_i\,; \ \sum_{i=1}^{\ell} \lambda_i y_i = 0; \quad Mi(w,b) \geq 1 - \xi_i \\ \xi_i \geq 0, \qquad \eta_i \geq 0, \qquad \lambda_i \geq 0, \qquad \eta_i + \lambda_i = C \\ \lambda_i M_i(w,b) = 0 \ \Longrightarrow \ \lambda_i = 0 \text{ або } M_i(w,b) = 1 - \xi_i, \qquad i = 1 \dots \ell \\ \eta_i \xi_i = 0 \ \Longrightarrow \ \eta_i = 0 \text{ або } \xi_i = 0, \qquad i = 1 \dots \ell \end{cases}$$

Діапазон значень λ_i які відповідають обмеженням на величину відступа дозволяють поділити об'єкти навчальної вибірки на три типи:

1. $\lambda_i = 0 \ \Longrightarrow \ \eta_i = C; \ \xi_i = 0; \ M_i \geq 1$ — периферійні (неінформативні) об'єкти: вони знаходяться в своєму класі, класифікуються вірно та не впливають на вибір розділяючої гіперплощини;
2. $0 < \lambda_i < C \ \Longrightarrow \ 0 < \eta_i < C; \ \xi_i = 0; \ M_i = 1$ — опорні граничні об'єкти: знаходяться чітко на границі розділяючої полоси на стороні свого класу;
3. $\lambda_i = C \ \Longrightarrow \ \eta_i = 0; \ \xi_i > 0; 0 < M_i < 1$ — об'єкти-порушники: знаходяться всередині розділяючої полоси;
4. $\lambda_i = C \ \Longrightarrow \ \eta_i = 0; \ \xi_i > 0; \ M_i < 0$ — помилки: не вірно класифіковані об'єкти, які знаходяться на боці не свого класу



Метод опорних векторів (SVM) в задачах класифікації

Перехід до більш зручного опису

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

	A		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	386.	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	60.07	244.	0

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

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

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

Відступ

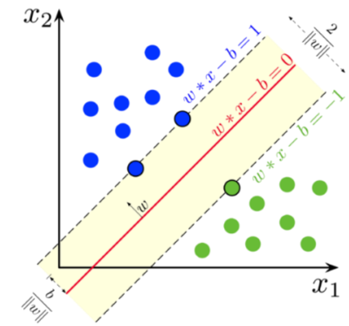
$$M_i(w) = yi\langle x_i, w\rangle$$

Задача оптимізації

$$\frac{1}{2}\|w\| + C\sum_{i=1}^{\ell}(1 - M_i(w))_+ \longrightarrow \min_w$$

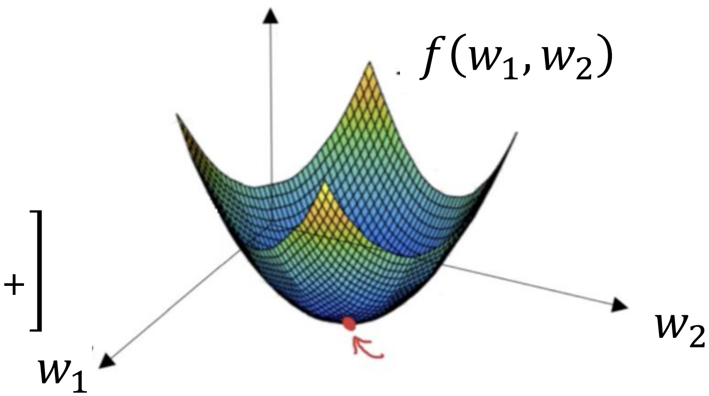
Розв'язується методом **градієнтного спуску**

Метод градієнтного спуску в задачах машинного навчання



Функціонал якості $f(w) \longrightarrow \min\limits_{w}$

$$f(w) = \frac{1}{2}\|w\| + C\frac{1}{\ell}\sum_{i=1}^{\ell}(1 - M_i(w))_+$$

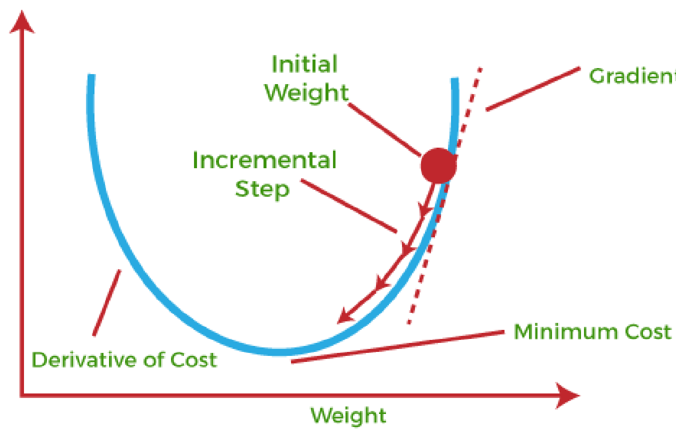


Для всіх об'єктів навчальної вибірки

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

Розрахунок наступного значення ваг w за методом градієнтного спуску:

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

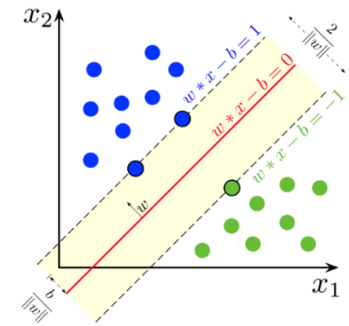


η – крок навчання

Похідна від функціоналу якості

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

Метод градієнтного спуску в задачах машинного навчання

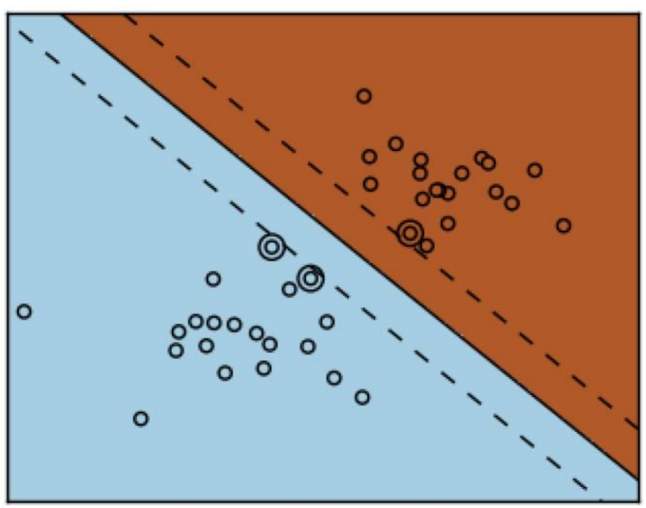


Розрахунок наступного значення ваг w
за методом градієнтного спуску

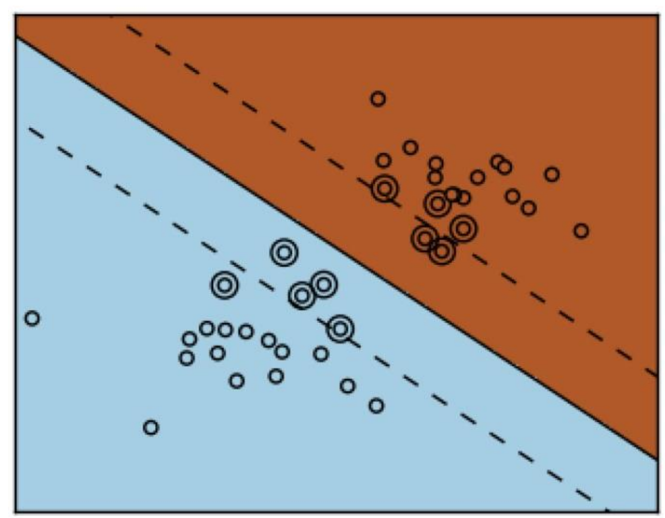
$$w_i = w_i - \eta \frac{1}{\ell} \sum_{i=1}^{\ell} \begin{cases} w, & if \max(1 - y_i \langle w, x_i \rangle)_+ = 0 \\ w - C y_i x_i, & otherwise \end{cases}$$

Вплив параметра C

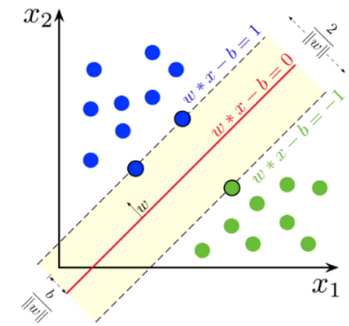
Велике значення параметра C
слабка регуляризація



Мале значення параметра C
сильна регуляризація



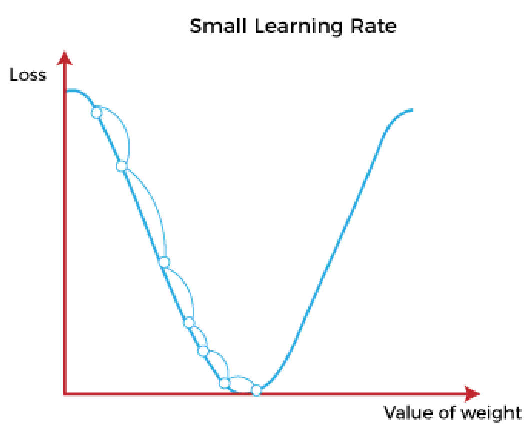
Метод градієнтного спуску в задачах машинного навчання

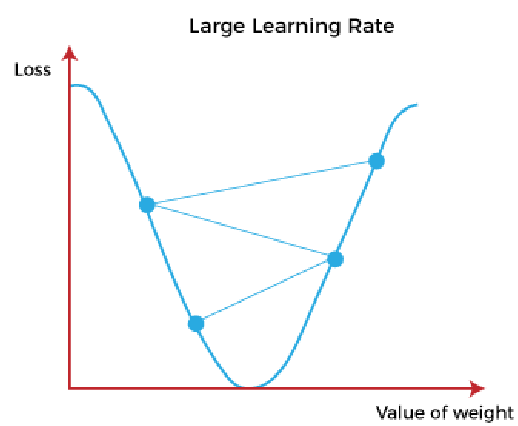


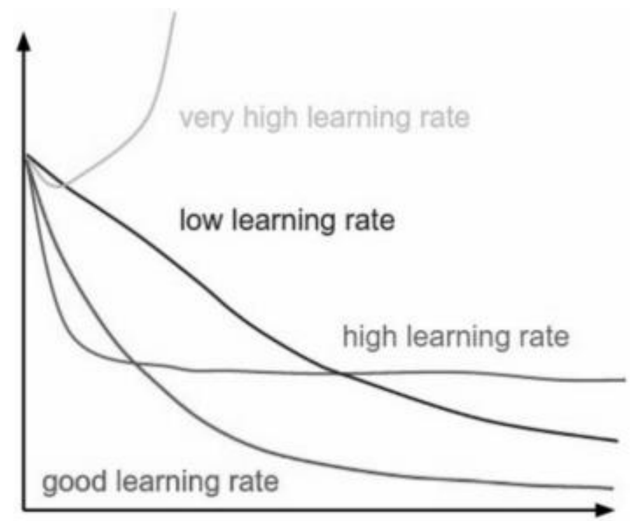
Розрахунок наступного значення ваг w за методом градієнтного спуску

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

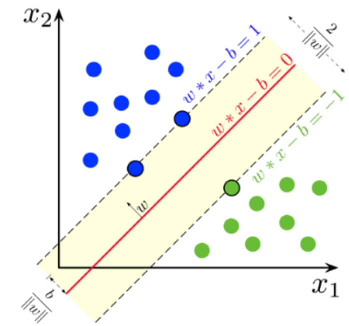
Вплив швидкості навчання η







Метод опорних векторів (SVM) в задачах класифікації



Нелінійне узагальнення (*kernel trick*)

- Для лінійно нероздільної вибірки $X = \mathbb{R}^n$ існує спрямляючий простір $\mathcal{H}$ (більшої розмірності) з функцією переходу $\psi: X \to \mathcal{H}$
- Скалярний добуток $\langle x_1, x_2 \rangle$ у просторі X замінінюється скалярним добутком $\langle \psi(x_1), \psi(x_2) \rangle$ у гільбертовому просторі $\mathcal{H}$ з визначеним скалярним добутком.
- Це надає можливість замінити скалярний добуток у просторі X на ядро — функцію, що є скалярним добутком у де-якому просторі $\mathcal{H}$. Замість підбору ψ можна підбирати безпосередньо ядро $K(x_i, x_j) = \langle \psi(x_1), \psi(x_2) \rangle$.

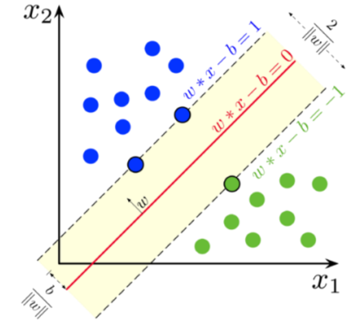
Постановка задачі з застосуванням ядер приймає вигляд

$$\begin{cases} -\mathcal{L}(\lambda) = -\sum_{i=1}^{\ell} \lambda_i + \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j K(x_i, x_j) \to \min\limits_{\lambda} \\ 0 \le \lambda_i \le C, \qquad i = 1 \ldots \ell \\ \sum_{i=1}^{\ell} \lambda_i y_i = 0 \end{cases}$$

Лінійний класифікатор з ознаками $f_i(x) = K(x, x_i)$:

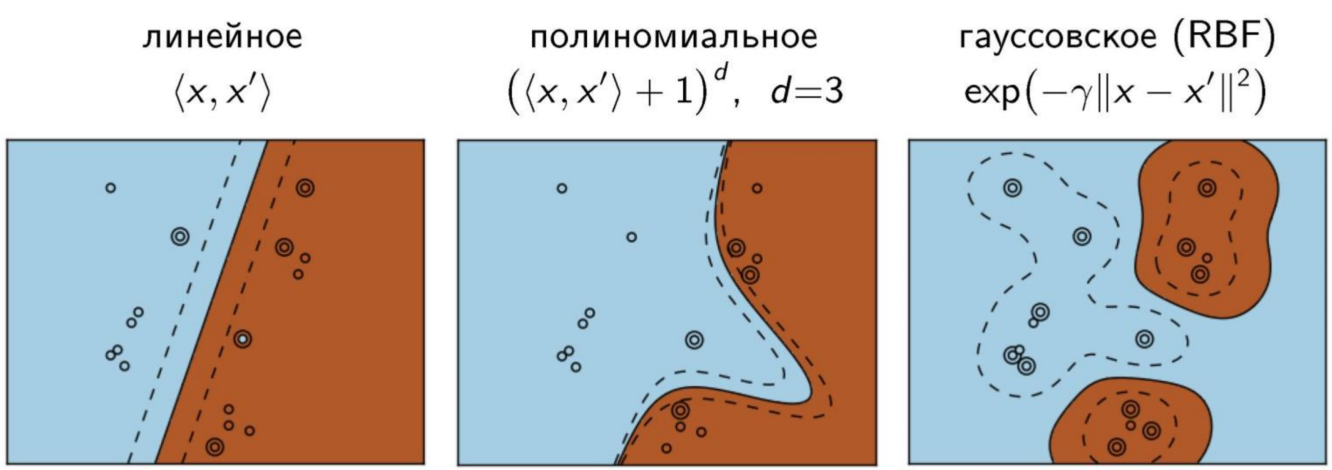
$$a(x) = sign\left(\sum_{i=1}^{\ell} \lambda_i y_i K(x, x_i) \right)$$

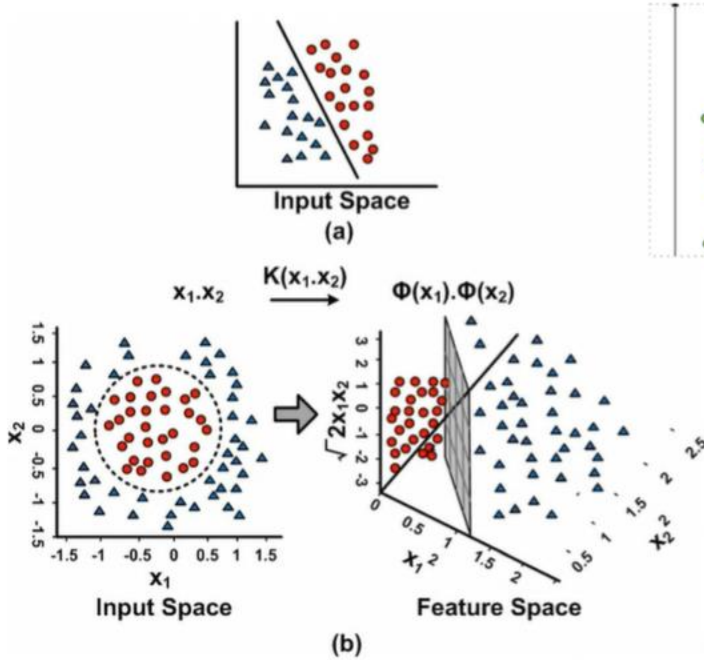
Метод опорних векторів (SVM) в задачах класифікації

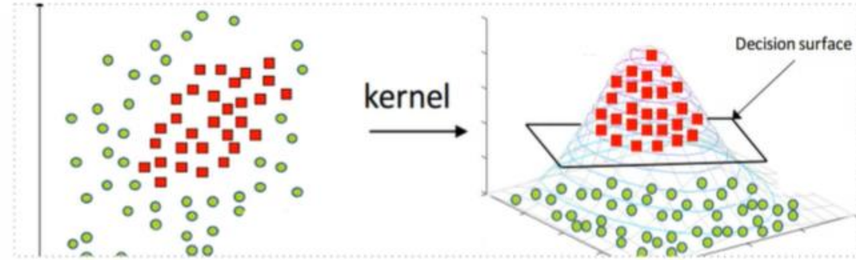


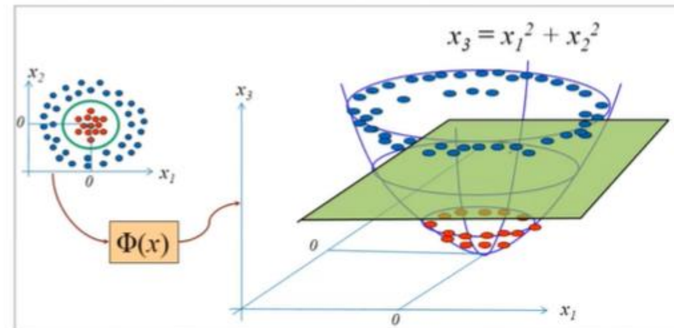
Приклади ядер

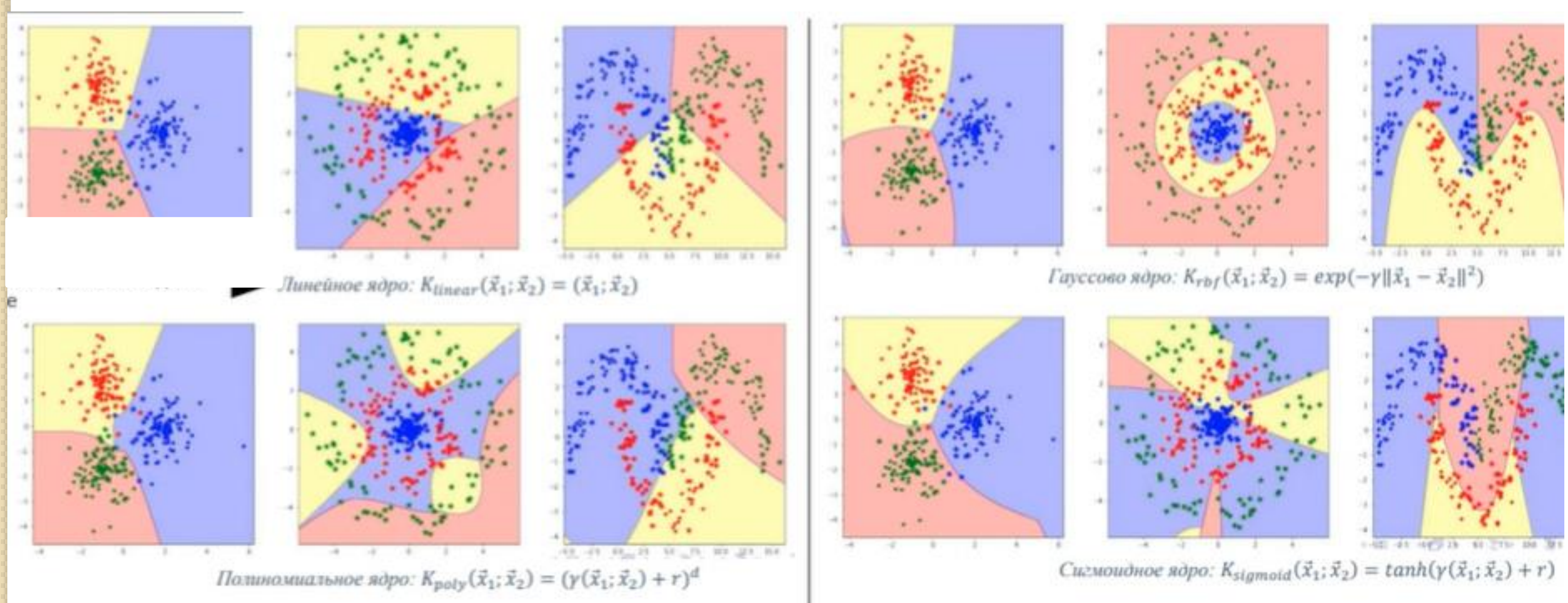
1. Квадратичне ядро $\dim \mathcal{H} = \frac{1}{2}n(n+1)$:
$$K(x, x') = \langle x, x' \rangle^2$$
2. Поліноміальне ядро, $\dim \mathcal{H} = C_{n+d-1}^{d}$:
$$K(x, x') = \langle x, x' \rangle^d$$
3. Поліноміальне ядро:
$$K(x, x') = (\langle x, x' \rangle + 1)^d$$
4. Сигноідне ядро:
$$K(x, x') = th(k_1 \langle x, x' \rangle - k_0), \qquad k_0, k_1 > 0$$
5. Гаусове ядро (RBF ядро):
$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$



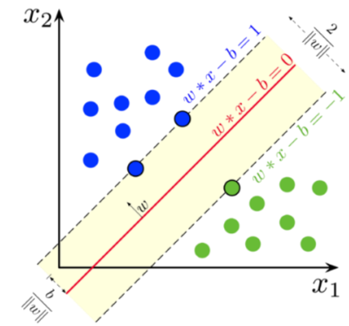

Input Space
(a)
$x_1.x_2$
$K(x_1.x_2)$
$\Phi(x_1).\Phi(x_2)$
Input Space
Feature Space
(b)


kernel
Decision surface


$x_3 = x_1^2 + x_2^2$
$\Phi(x)$


Линейное ядро: $K_{linear}(\vec{x}_1; \vec{x}_2) = (\vec{x}_1; \vec{x}_2)$
Гауссово ядро: $K_{rbf}(\vec{x}_1; \vec{x}_2) = exp(-\gamma\|\vec{x}_1 - \vec{x}_2\|^2)$
Полиномиальное ядро: $K_{poly}(\vec{x}_1; \vec{x}_2) = (\gamma(\vec{x}_1; \vec{x}_2) + r)^d$
Сигмоидное ядро: $K_{sigmoid}(\vec{x}_1; \vec{x}_2) = tanh(\gamma(\vec{x}_1; \vec{x}_2) + r)$

Метод опорних векторів (SVM) в задачах класифікації



Переваги та недоліки SVM

Переваги класичного SVM :

- Завдання опуклого квадратичного програмування добре вивчене і має єдине рішення.
- Метод опорних векторів еквівалентний двошарової нейронної 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.

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.

Дякую за увагу