

ЛЕКЦІЯ №4

Розділяючі поверхні. Модель нейрона

«Розділяючі поверхні. Модель нейрона»

Перед тим, як перейти до поняття розділяючих поверхонь в просторі ознак, проаналізуємо апостеріорні розподіли основані на Байєсовському підході.

Класифікатор Байєса.

Якщо вибірка випадкових подій n_0 за несумісними гіпотезами $H = (h_1, h_2, \dots, h_\lambda, \dots, h_\Lambda)$, де $\lambda = 1, 2, \dots, \Lambda$ – порядковий номер гіпотези, розділена на відповідні окремі групи $n_{h\lambda}$, то з неї можна отримати формулу для повної групи ймовірностей, підсумок яких дорівнює 1.

$$n_0 = \sum_{\lambda=1}^{\Lambda} n_{h\lambda}; \quad 1 = \sum_{\lambda=1}^{\Lambda} n_{h\lambda} / n_0; \quad 1 = \sum_{\lambda=1}^{\Lambda} p_{h\lambda}, \quad (5.4.1)$$

де $p_{h\lambda} = n_{h\lambda} / n_0$ – ймовірність гіпотези h_λ , $\lambda = 1, 2, \dots, \Lambda$ – порядковий номер гіпотези.

Якщо кожна група $n_{h\lambda}$ має деяку власну підгрупу $n_{A\lambda}$ з випадковими подіями A_λ , то отримуємо аналогічні вирази для повної групи умовних ймовірностей

$$n_A = \sum_{\lambda=1}^{\Lambda} n_{A\lambda}; 1 = \sum_{\lambda=1}^{\Lambda} n_{A\lambda} / n_A; 1 = \sum_{\lambda=1}^{\Lambda} p(A_\lambda | A) \quad (5.4.2)$$

де $p(A_\lambda|A) = n_{A\lambda} / n_A = p_{A\lambda}$ представляє умовну ймовірність: «ймовірність спостереження подій A_λ , якщо всі події A відбулись». Вертикальна риска у позначенні $p(A_\lambda|A)$ відноситься до подій A і тлумачиться словом «якщо». З формул (5.4.2) отримуємо формули повної ймовірності:

$$n_A = \sum_{\lambda=1}^{\Lambda} n_{A\lambda}; p(A) = \frac{n_A}{n_0} = \sum_{\lambda=1}^{\Lambda} \frac{n_{A\lambda}}{n_0} = \sum_{\lambda=1}^{\Lambda} \left(\frac{n_{A\lambda}}{n_A} \cdot \frac{n_A}{n_0} \right) = \sum_{\lambda=1}^{\Lambda} [p(A_\lambda | A)p(A)]; \quad (5.4.3)$$

$$p(A) = \frac{n_A}{n_0} = \sum_{\lambda=1}^{\Lambda} \frac{n_{A\lambda}}{n_0} = \sum_{\lambda=1}^{\Lambda} \left(\frac{n_{A\lambda}}{n_{h\lambda}} \cdot \frac{n_{h\lambda}}{n_0} \right) = \sum_{h=1}^H p(A_\lambda | h_\lambda) p_{h\lambda} \quad (5.4.4)$$

Припустимо, що у вибірці спостережених вогнищ n_0 ймовірності спостереження окремих подій A (полум'я), B (дим) та їх сумісного спостереження AB (полум'я та дим) дорівнюють

$$p(A) = \frac{n_A}{n_0}; \quad p(B) = \frac{n_B}{n_0}; \quad p(AB) = \frac{n_{AB}}{n_0}, \quad (5.4.5)$$

де $n_A = n_{A1} + n_{AB}$ – кількість спостережених подій A (полум'я); $n_B = n_{B1} + n_{AB}$ – кількість спостережених подій B (дим); n_{AB} – кількість одночасно спостережених подій (полум'я та дим), яка одночасно входить у множини n_A та n_B .

Умовні ймовірності спостереження події A (якщо події B вже відбулись) та події B (якщо події A вже відбулись) дорівнюють

$$p(A|B) = \frac{n_{AB}}{n_B}, \quad p(B|A) = \frac{n_{AB}}{n_A}. \quad (5.4.6)$$

Ймовірність одночасного спостереження двох подій дорівнює

$$p(AB) = \frac{n_{AB}}{n_0} = \frac{n_{AB}}{n_B} \cdot \frac{n_B}{n_0} = p(A|B)p(B); \quad (5.4.7)$$

$$p(AB) = \frac{n_{AB}}{n_0} = \frac{n_{AB}}{n_A} \cdot \frac{n_A}{n_0} = p(B | A)p(A). \quad (5.4.8)$$

Якщо прирівняти (5.4.7) та (5.4.8), то отримуємо правило Байєса

$$p(A | B) = \frac{p(B | A)p(A)}{p(B)}, \quad p(B | A) = \frac{p(A | B)p(B)}{p(A)}. \quad (5.4.9)$$

Напрямки аналізу ймовірнісних мереж за методом Байєса. Для розпізнавання кожної групи – гіпотези вектора $H = (h_1, h_2, \dots, h_\lambda, \dots, h_\Lambda)$, де $\lambda = 1, 2, \dots, \Lambda$ – порядковий номер гіпотези початкової вибірки об'єктів n_0 , використовують ознаки – змінні об'єктів вибірки $X = (x_1, x_2, \dots, x_j, \dots, x_n)$, кожна з яких для окремого об'єкта приймає одне з двох значень: «0 /+ 1», «-1/+1», «Орел / Решка» тощо.

У залежності від кількості ознак-змінних задачі Байєса можна розділити на дві групи:

- вибірка з однією ознакою x_1 ;
- вибірка з кількома ознаками $x_j, j = 1, \dots, n$.

Кожна з цих вибірок може мати задану кількість взаємно несумісних гіпотез $H = (h_1, h_2, \dots, h_\lambda, \dots, h_\Lambda)$, де $\lambda = 1, 2, \dots, \Lambda$ – порядковий номер гіпотези, ймовірності яких $p_{h\lambda}$ складають повну групу.

1. Вибірка з об'єктами, що описуються однією ознакою x_1 .

Одна ознака – змінна x_1 може приймати одне з двох значень: або x_{11} ; $h_\lambda=1$, або x_{10} ; $h_\lambda=0$. Тому кожна група гіпотези розділяється на дві підгрупи у вигляді $[(n_{h_1}; x_{11}, n_{h_1}; x_{10}), \dots, (n_{h_\lambda}; x_{11}, n_{h_\lambda}; x_{10}), \dots, (n_{h_\Lambda}; x_{11}, n_{h_\Lambda}; x_{10})]$ з дотриманням рівності $n_{h\lambda} = n_{h\lambda}; x_{11} + n_{h\lambda}; x_{10}$.

Ми можемо:

- Визначити апріорну ймовірність спостереження гіпотези h_λ , $\lambda = 1, 2, \dots, \Lambda$, за формулою:

$$P_{h\lambda, x_{11}} = \frac{n_{h\lambda; x_{11}}}{n_0} = \frac{n_{h\lambda}}{n_0} \cdot \frac{n_{h\lambda; x_{11}}}{n_{h\lambda}} = P_{h\lambda} p(x_{11}; h\lambda | h_\lambda), \quad (5.4.11)$$

де $p_{h\lambda} = n_{h\lambda} / n_0$ – априорна ймовірність гіпотези $h\lambda$; $p(x_{11}; h\lambda | h_\lambda) = n_{h\lambda; x_{11}} / n_{h\lambda}$ – умовна ймовірність спостереження ознаки $x_{11}; h\lambda=1$ в групі $n_{h\lambda}$ (гіпотези $h\lambda$); $p(x_{10}; h\lambda | h_\lambda) = 1 - p(x_{11}; h\lambda | h_\lambda)$ – умовна ймовірність спостереження ознаки $x_{10}; h\lambda=0$ (або $x_{10}; h\lambda=-1$) в групі $n_{h\lambda}$.

– Розрахувати ймовірність спостереження об'єктів з заданим значенням ознаки – змінної $x_{11}; h\lambda=+1$ для всієї вибірки

$$P_{x_{11}} = \sum_{\lambda=1}^{\Lambda} \frac{n_{h\lambda; x_{11}}}{n_0} = \sum_{\lambda=1}^{\Lambda} \left(\frac{n_{h\lambda}}{n_0} \cdot \frac{n_{h\lambda; x_{11}}}{n_{h\lambda}} \right) = \sum_{\lambda=1}^{\Lambda} P_{h\lambda} p(x_{11}; h\lambda | h_\lambda). \quad (5.4.12)$$

З формул (5.4.11) та (5.4.12) визначаємо формулу Байєса для однієї змінної при наявності кількох гіпотез, за якою ймовірність спостереження гіпотези $h\lambda$ при умові появи ознаки $x_{11}; h\lambda=+1$ дорівнює

$$p(h_\lambda | x_{11}) = \frac{n_{h\lambda, x_{11}}}{\sum_{\lambda=1}^{\Lambda} n_{h\lambda, x_{11}}} = \frac{n_{h\lambda, x_{11}}}{n_0 \frac{1}{\sum_{\lambda=1}^{\Lambda} n_{h\lambda, x_{11}}}} = \frac{P_{h\lambda, x_{11}}}{P_{x_{11}}} = \frac{P_{h\lambda} p(x_{11}; h\lambda | h_\lambda)}{\sum_{\lambda=1}^{\Lambda} P_{h\lambda} p(x_{11}; h\lambda | h_\lambda)}. \quad (5.4.13)$$

2. У вибірці з об'єктами, що описуються багатьма змінними $X = (x_1, x_2, \dots, x_j, \dots, x_n)$, при умові незалежності змінних, елементи вектора X для однієї гіпотези спостерігаються одночасно. Тому в формулі (5.4.13) ймовірність спостереження однієї гіпотези h_λ замінюється добутком ймовірностей спостереження кількох ознак – змінних вектора X . В результаті отримуємо відому формулу наївного Байєсового класифікатора при спостереженні подій – гіпотез (об'єктів) багатомірного простору, за якою ймовірність спостереження гіпотези h_λ при класифікації невідомого вектора X

$$P_{h_\lambda}^D = \frac{P_{h_\lambda} \prod_{j=1}^n p(x_j | h_\lambda)}{\sum_{\lambda=1}^L [P_{h_\lambda} \prod_{j=1}^n p(x_j | h_\lambda)]}, \quad (5.4.14)$$

де $p_{h\lambda} = n_{h\lambda}/n_0$ – апіорна ймовірність гіпотези $h\lambda$; $p(x_j|h\lambda)$ – ймовірність наявності ознаки x_j в гіпотезі $h\lambda$.

Тому що в навчальній групі $n_{h\lambda}$ гіпотези $h\lambda$ існують навчальні об'єкти з ознаками $x_j = +1$ та $x_j = -1$, то, внаслідок їх рівнозначності при розпізнаванні образів, при навчанні класифікатора Байєса в окремій таблиці визначають обидві значення ймовірностей у вигляді $p(x_j|h\lambda)$ та $p^*(x_j|h\lambda)$.

Для конкретного невідомого об'єкта X , який класифікується, використовують з цієї таблиці одну з умовних ймовірностей у залежності від знаку x_j об'єкта класифікації X : або $p(x_j|h\lambda)$ або $p^*(x_j|h\lambda)$. Якщо при навчанні отримано $p(x_j|h\lambda) = 0$ або $p^*(x_j|h\lambda)=0$, то такі значення перетворюються, наприклад, за методом згладжування Лапласа на «приблизний нуль» (див. далі), тому що ймовірність спостереження існуючої гіпотези $p_{h\lambda}$ не може дорівнювати нулю.

Формула (5.4.14) широко застосовується і підтверджується практикою, але вона має наступний недолік, який збільшує похибку класифікації гіпотези: якщо $p(x_j|h_\lambda)$ дорівнює нулю, або близька до нуля, то порушується основний принцип статистики: одноосібне «рішення однієї змінної x_j » у виразі $\prod_{j=1}^n p(x_j | h_\lambda)$ «забороняє» «рішення» всіх інших змінних, навіть якщо всі їх умовні ймовірності $p(x_j|h_\lambda)$ дорівнюють 1 (тобто якщо всі інші змінні свідчать на користь прийняття гіпотези).

Цього недоліку позбавлена формули Байєсового класифікатора, які рекомендуються до використання

$$P_{h\lambda}^{\Sigma} = \frac{P_{h\lambda} \sum_{j=1}^n p(x_j | h_\lambda)}{\sum_{\lambda=1}^{\Lambda} [P_{h\lambda} \sum_{j=1}^n p(x_j | h_\lambda)]}, \quad (5.4.15)$$

$$P_{h\lambda}^{p\Sigma} = P_{h\lambda} \sum_{j=1}^n p(x_j | h_\lambda), \quad (5.4.16)$$

Формули Байєсового класифікатора (5.4.15) та(5.4.16), на відміну від формули (5.4.14), отримані на основі принципу математичного очікування ознак гіпотези, що замінює добуток у формулі (5.4.14) на підсумок. В результаті усуваються відповідні стохастичні похибки класифікації формули (5.4.14): відпадає ніколи не дотримувана і невірна вимога взаємної незалежності ознак x_j (ця незалежність математично означає неможливість розпізнати окремих об'єкт, який повинен мати індивідуальні залежні між собою ознаки); жодна змінна x_j не має переваг перед іншими змінними; цілком природна і нормальна нульова вхідна інформація не псується шляхом згладжування (надання їй не спостереженого числового значення); ліквідується порушення основного принципу статистики, коли одна ознака з малим значенням ймовірності «одноосібно забороняє» обрання гіпотези всупереч протилежному «колективному рішенню» всіх інших ознак.

Задача розпізнавання.

В загальному формулюванні, задача розпізнавання – віднесення представленого об'єкту до того чи іншого класу на підставі його дослідження у просторі визначених ознак.

При цьому виникає ризик не правильної класифікації. Відповідно актуальним є питання зменшення ймовірності помилки класифікації.

Дискримінатні функції і поверхні рішень.

Мінімізація ризику і ймовірність помилки еквівалентні розділенню простору ознак на M областей. Якщо області R_i і R_j суміжні, то вони розділені поверхнею рішень в багатовимірному просторі. Для випадку мінімізації ймовірності похибки, поверхня рішень задається рівнянням:

$$P(\Omega_i | x) - P(\Omega_j | x) = 0$$

В даному рівнянні, доводиться працювати з ймовірностями. Часто замість ймовірностей вигідно працювати з деякими функціями від ймовірностей. При складних перетвореннях, які вміщують добуток ймовірностей, доцільно перейти до логарифмів складових.

$$g_i(x) = f(P(\Omega_i | x)),$$

де функція f – монотонно зростає.

Визначення.

Функція $g_i(x) = f(P(\Omega_i | x))$ називається дискримінантна функція.

Таким чином, поверхня рішень, буде задаватися рівнянням:

$$g_i(x) - g_j(x) = 0, \quad i = 1, 2 \dots M, \quad j = 1, 2 \dots M, \quad i \neq j.$$

Для задачі класифікації по ймовірності похибки чи ризику не завжди вдається обчислити ймовірності. В такому випадку, доцільно розрахувати розділяючу поверхню на основі іншої функції.

Байєсовський класифікатор для нормального розподілу.

Розподіл Гауса, широко застосовується із - за комфортності розрахунків і адекватності в багатьох випадках (чому?). Розглянемо багатовимірну щільність нормального розподілу $N(\mu_i, \Sigma_i)$:

$$p(x|\Omega_i) = \frac{1}{(2\pi)^{l/2} |\Sigma_i|^{1/2}} \cdot \exp\left(-\frac{1}{2} \frac{(x - \mu_i)^T}{\Sigma_i \cdot (x - \mu_i)}\right), \quad i = 1, 2, \dots, M$$

де $\mu_i = E[X]$ - математичне сподівання випадкової величини x в класі Ω_i ,

Σ_i - матриця коваріації розмірності $l \times l$ для класу Ω_i , $\Sigma_i = E[(x - \mu_i) \cdot (x - \mu_i)^T]$,

$|\Sigma_i|$ - визначник матриці коваріації.

Тут x, μ_i - це вектори – стовбчики, а x^T, μ_i^T - вектори – рядки.

Квадратична поверхня рішень.

На підставі цих даних, необхідно побудувати байєсовський класифікатор. Розглянемо логарифмічну дискримінантну функцію:

$$\begin{aligned}g_i(x) &= \ln(P(\Omega_i|x)) = \\&= \ln(p(x|\Omega_i)P(\Omega_i)) = \\&= \ln p(x|\Omega_i) + \ln P(\Omega_i) = \\&= -\frac{1}{2} \frac{(x - \mu_i)}{\Sigma_i} (x - \mu_i)^T + \ln P(\Omega_i) + \ln \frac{1}{(2\pi)^{l/2} |\Sigma_i|^{1/2}} = \\&= -\frac{1}{2} \frac{(x - \mu_i)}{\Sigma_i} (x - \mu_i)^T + \ln P(\Omega_i) - \frac{l}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| = \\&= -\frac{1}{2} \frac{(x - \mu_i)}{\Sigma_i} (x - \mu_i)^T + \ln P(\Omega_i) + C_i, \text{ где } C_i = -\frac{l}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i|\end{aligned}$$

Ця функція, являє собою квадратичну форму. Значить розділяюча поверхня $g_i(x) - g_j(x) = 0$ являє собою гіперповерхню другого порядку. Тому Байєсовський класифікатор є квадратичним класифікатором.

Приклад №1:

Нехай $l = 2$, $\Sigma_i = \begin{pmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_i^2 \end{pmatrix}$. Тоді $\frac{1}{\Sigma_i} = \begin{pmatrix} \frac{1}{\sigma_i^2} & 0 \\ 0 & \frac{1}{\sigma_i^2} \end{pmatrix}$.

$$g_i(x) = -\frac{1}{2\sigma_i^2}(x_1^2 + x_2^2) + \frac{1}{\sigma_i^2}(\mu_{i1}x_1 + \mu_{i2}x_2) - \frac{1}{2\sigma_i^2}(\mu_{i1}^2 + \mu_{i2}^2) + \ln(P(\Omega_i)) + C_i$$

Розділяючою поверхнею є конічний переріз.

Приклад №2:

Нехай $P(\Omega_1) = P(\Omega_2)$, $\mu_1 = (0,0)$, $\mu_2 = (1,0)$, $\Sigma_1 = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.15 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.25 \end{pmatrix}$.

Тоді $\frac{1}{\Sigma_1} = \begin{pmatrix} 10 & 0 \\ 0 & \frac{20}{3} \end{pmatrix}$, $\frac{1}{\Sigma_2} = \begin{pmatrix} 5 & 0 \\ 0 & 4 \end{pmatrix}$. Знайдемо поверхню рішення.

$$g_1(x) = -\frac{1}{2}(x_1, x_2) \begin{pmatrix} 10 & 0 \\ 0 & \frac{20}{3} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \ln P(\Omega_1) - \ln(2\pi) + \frac{1}{2} \ln \frac{200}{3} =$$

$$= -\frac{1}{2} \left(10x_1^2 + \frac{20}{3}x_2^2 \right) + \ln P(\Omega_1) - \ln(2\pi) + \frac{1}{2} \ln \frac{200}{3}$$

$$g_2(x) = -\frac{1}{2}(x_1 - 1, x_2) \begin{pmatrix} 5 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix} + \ln P(\Omega_2) - \ln(2\pi) + \frac{1}{2} \ln 20 =$$

$$= -\frac{1}{2} (5(x_1 - 1)^2 + 4x_2^2) + \ln P(\Omega_2) - \ln(2\pi) + \frac{1}{2} \ln 20$$

$$g_1(x) - g_2(x) = -\frac{1}{2} \left(10x_1^2 + \frac{20}{3}x_2^2 - 5(x_1 - 1)^2 - 4x_2^2 \right) + \frac{1}{2} \left(\ln \frac{200}{3} - \ln 20 \right) =$$

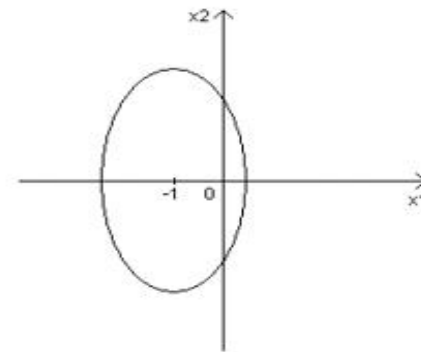
$$= -\frac{1}{2} \left(5(x_1 + 1)^2 + \frac{8}{3}x_2^2 \right) + 5 + \frac{1}{2} \ln \frac{10}{3}$$

Т.к. $g_1(x) - g_2(x) = 0$, то $-\frac{1}{2} \left(5(x_1 + 1)^2 + \frac{8}{3}x_2^2 \right) + 5 + \frac{1}{2} \ln \frac{10}{3} = 0$

$$5(x_1 + 1)^2 + \frac{8}{3}x_2^2 = 10 + \ln \frac{10}{3}$$

$$\frac{(x_1 + 1)^2}{\frac{8}{3}} + \frac{x_2^2}{5} = \frac{3}{40} \left(10 + \ln \frac{10}{3} \right)$$

$$\frac{(x_1 + 1)^2}{\left(2\sqrt{\frac{2}{3}}\right)^2} + \frac{x_2^2}{(\sqrt{5})^2} = \frac{3}{40} \left(10 + \ln \frac{10}{3} \right)$$



Значить рівняння поверхні рішень – еліпс з центром в точці $(-1; 0)$.

Приклад №3:

Нехай $P(\Omega_1) = P(\Omega_2)$, $\mu_1 = (0,0)$, $\mu_2 = (1,0)$, $\Sigma_1 = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.15 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.25 \end{pmatrix}$.

Тоді $\frac{1}{\Sigma_1} = \begin{pmatrix} 10 & 0 \\ 0 & \frac{20}{3} \end{pmatrix}$, $\frac{1}{\Sigma_2} = \begin{pmatrix} 5 & 0 \\ 0 & 4 \end{pmatrix}$. Знайдемо поверхню рішення.

З попереднього прикладу №2, отримаємо:

$$g_1(x) = -\frac{1}{2}(5(x_1 - 1)^2 + 4x_2^2) + \ln P(\Omega_2) - \ln(2\pi) + \frac{1}{2} \ln 20$$

$$g_2(x) = -\frac{1}{2}(x_1 - 1, x_2) \begin{pmatrix} \frac{20}{3} & 0 \\ 0 & 10 \end{pmatrix} \begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix} + \ln P(\Omega_2) + \frac{1}{2} \ln \frac{200}{3}$$

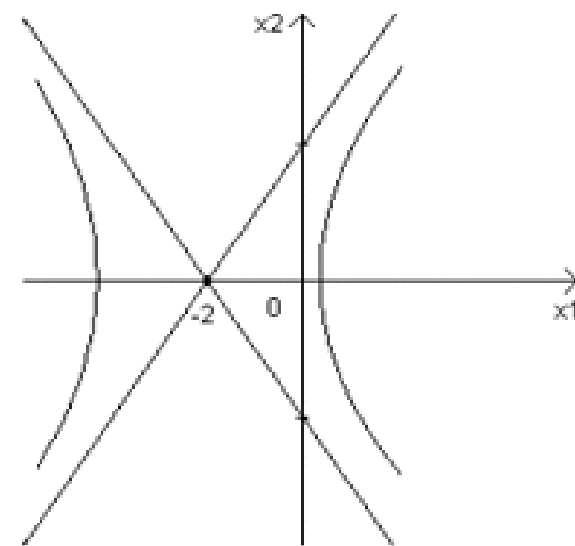
$$g_1(x) - g_2(x) = -\frac{1}{2} \left(10x_1^2 + \frac{20}{3}x_2^2 - \frac{20}{3}(x_1 - 1)^2 - 10x_2^2 \right) =$$

$$= -\frac{1}{2} \left(\frac{10}{3}x_1^2 - \frac{10}{3}x_2^2 + \frac{40}{3}x_1 - \frac{20}{3} \right) =$$

$$= -\frac{1}{2} \cdot \frac{10}{3} (x_1^2 - x_2^2 + 4x_1 - 2) = -\frac{5}{3} ((x_1 + 2)^2 - x_2^2 - 6)$$

Т.к. $g_1(x) - g_2(x) = 0$, то $-\frac{5}{3} ((x_1 + 2)^2 - x_2^2 - 6) = 0$.

$(x_1 + 2)^2 - x_2^2 = 6$. Гіпербола з центром в точці $(-2; 0)$.



Лінійна поверхня рішення.

Для тієї ж умови:

$$p(x|\Omega_i) = \frac{1}{(2\pi)^{l/2} \cdot |\Sigma_i|} \cdot \exp\left(-\frac{1}{2} \frac{x - \mu_i}{\Sigma_i} (x - \mu_i)^T\right), \quad i = 1, 2, \dots, M.$$

Вище, ми отримали квадратичну форму:

$$\begin{aligned} g_i(x) &= \ln(P(\Omega_i|x)) = \\ &= \ln(p(x|\Omega_i)P(\Omega_i)) = \\ &= \ln p(x|\Omega_i) + \ln P(\Omega_i) = \\ &= -\frac{1}{2} \frac{(x - \mu_i)}{\Sigma_i} (x - \mu_i)^T + \ln P(\Omega_i) + \ln \frac{1}{(2\pi)^{l/2} |\Sigma_i|^{1/2}} = \\ &= -\frac{1}{2} \frac{(x - \mu_i)}{\Sigma_i} (x - \mu_i)^T + \ln P(\Omega_i) - \frac{l}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| = \\ &= -\frac{1}{2} \frac{(x - \mu_i)}{\Sigma_i} (x - \mu_i)^T + \ln P(\Omega_i) + C_i, \quad \text{где } C_i = -\frac{l}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| \end{aligned}$$

Нехай $\Sigma_i = \Sigma_j$, тоді:

$$\begin{aligned}h_i(x) &= -\frac{1}{2} \left[\frac{x}{\Sigma_i} x^T - \frac{\mu_i}{\Sigma_i} x^T - \frac{x}{\Sigma_i} \mu_i^T + \frac{\mu_i}{\Sigma_i} \mu_i^T \right] + \ln P(\Omega_i) + C_i = \\&= -\frac{1}{2} \left[\frac{x}{\Sigma_i} x^T - 2 \frac{\mu_i}{\Sigma_i} x^T + \frac{\mu_i}{\Sigma_i} \mu_i^T \right] + \ln P(\Omega_i) + C_i = \\&= -\frac{1}{2} \left[K_i(x) - 2W_i x^T + W_i \mu_i^T \right] + \ln P(\Omega_i) + C_i = \\&= -\frac{1}{2} K_i(x) + L_i(x) + C_i, \text{ где } L_i(x) = W_i x^T + W_{i0}; W_i = \frac{\mu_i}{\Sigma_i};\end{aligned}$$

$$W_{i0} = \ln P(\Omega_i) - \frac{1}{2} W_i \mu_i^T$$

При $\Sigma_i = \Sigma_j$ порівнювати можна тільки $L_i(x)$ і $L_j(x)$. Таким чином, при $\Sigma_i = \Sigma_j$, отримаємо **лінійну поверхню рішень**.

Лінійна поверхня рішень з діагональною матрицею коваріації

Розглянемо випадок, коли матриця Σ діагональна з однаковими елементами:

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}. \text{Тоді } L_i(x) \text{ матиме вигляд:}$$

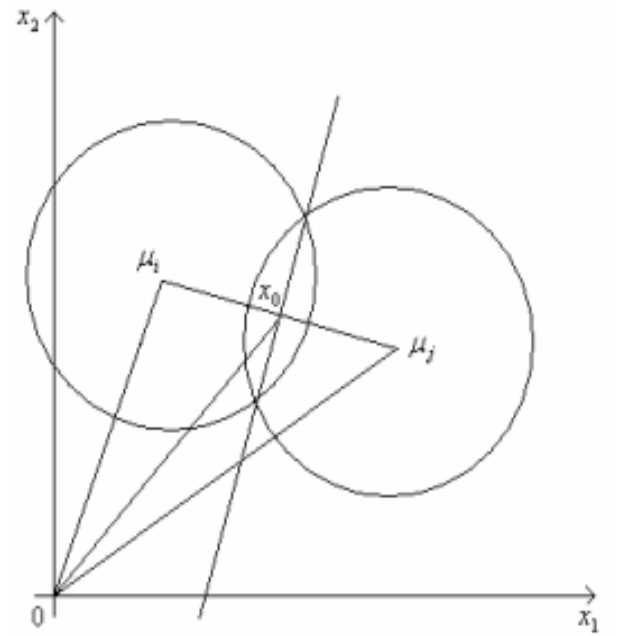
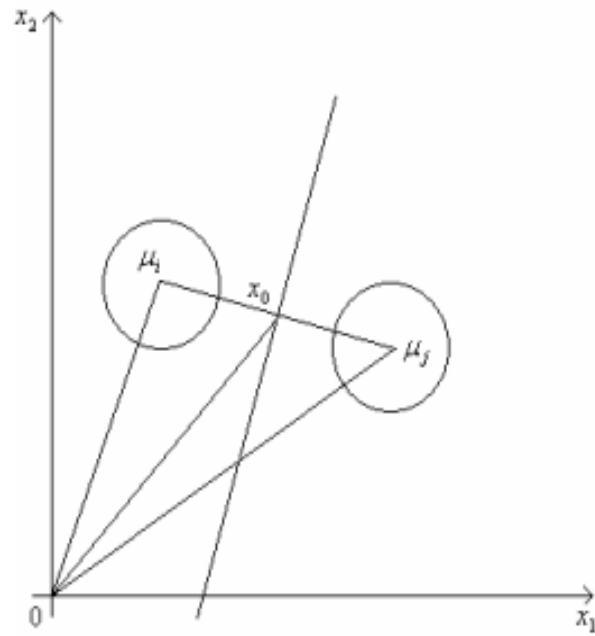
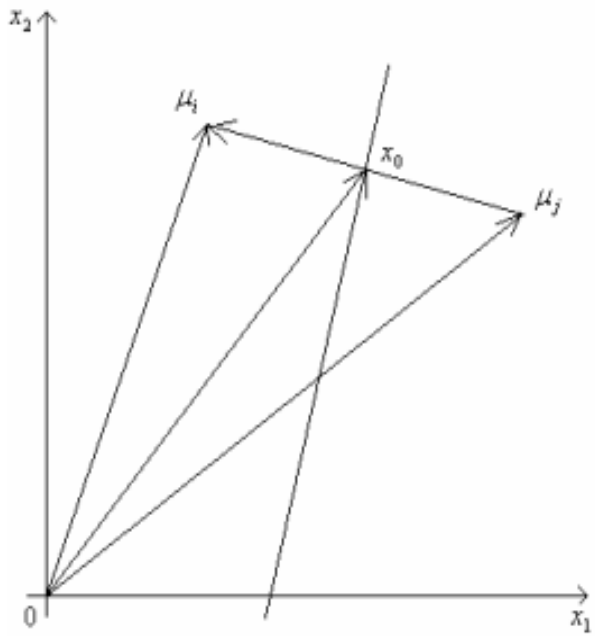
$$L_i(x) = \frac{1}{\sigma^2} \mu_i^T x + W_{i0}; \quad L_{ij}(x) = L_i(x) - L_j(x) = W^T (x - x_0) = 0,$$

де $W = \mu_i - \mu_j$, $x_0 = \frac{1}{2}(\mu_i + \mu_j) - \sigma^2 \frac{\mu_i - \mu_j}{\|\mu_i - \mu_j\|^2} \ln\left(\frac{P(\Omega_i)}{P(\Omega_j)}\right)$. В даному випадку під

нормою розуміється евклідова норма. Поверхнею рішення, є **гіперплощина**, яка проходить через точку x_0 .

Якщо $P(\Omega_1) = P(\Omega_2)$, то x_0 середина вектора $\overline{\mu_i \mu_j}$. Оскільки $L_{ij}(x) = 0$, то

$W^T (x - x_0) = (\mu_i - \mu_j)^T (x - x_0) = 0$. Значить поверхня рішення ортогональна вектору $\overline{\mu_i \mu_j}$. Схеми нижче демонструють різні положення лінійної розділяючої поверхні.



Приклад №4:

Приклад розділяючої поверхні рішення для двокласової задачі з нормальним розподілом.

Поверхня рішень лежить ближче до μ_i , якщо $P(\Omega_1) < P(\Omega_2)$. Відповідно, поверхня рішень лежить ближче до μ_j , якщо $P(\Omega_1) > P(\Omega_2)$. Також, якщо σ^2 не велике по відношенню до $\|\mu_i - \mu_j\|$ то розташування поверхні рішення не занадто чутливе до зміни $P(\Omega_1)$, $P(\Omega_2)$. Останнє твердження очевидне, оскільки вектори лежать в малих околицях μ_i , μ_j . На схемі вище, в центрі зображено випадок малого, а праворуч великого σ^2 .

Лінійна розділяюча поверхня з недіагональною матрицею коваріації.

В такому випадку рівняння

$$L_i(x) = \frac{1}{\sigma^2} \mu_i^T x + W_{i0}; \quad L_{\bar{i}}(x) = L_i(x) - L_j(x) = W^T (x - x_0) = 0, \quad \text{буде мати дещо інші}$$

параметри $W = \frac{\mu_i - \mu_j}{\Sigma}$; $x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\mu_i - \mu_j}{\|\mu_i - \mu_j\|_{\Sigma^{-1}}^2}$. В даному випадку, під нормою

вважається Σ^{-1} норма вектора \vec{a} , яка має вигляд $\|\vec{a}\|_{\Sigma^{-1}} = (\vec{a}^T \Sigma^{-1} \vec{a})$. Для такої норми, поверхня рішення не ортогональна вектору $\overline{\mu_i \mu_j}$, зате ортогональна його образу при перетворенні $\Sigma^{-1}(\mu_i - \mu_j)$.

Класифікатори по мінімуму відстані.

Будемо розглядати рівноймовірнісні класи з однаковою матрицею коваріації. Тоді

$\Sigma_1 = \Sigma_2 = \dots = \Sigma_n = \Sigma$ і вираз $L_i(x) = -\frac{1}{2}(x - \mu_i)\Sigma_i^{-1}(x - \mu_i)^T + \ln(P(\Omega_i)) + C_i$ набуде

вигляду (логарифм і константа скоротяться)

$$L_i(x) = -\frac{1}{2}(x - \mu_i)\Sigma_i^{-1}(x - \mu_i)^T.$$

Класифікатори по мінімуму відстані з діагональною матрицею коваріації.

Розглянемо випадок, коли матриця Σ діагональна з однаковими елементами:

$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. Тоді максимізація $L_i(x) = -\frac{1}{2}(x - \mu_i)\Sigma_i^{-1}(x - \mu_i)^T$ забезпечує

мінімізацію евклідової відстані $d_E = \|x - \mu_i\|$. При цьому вважається, що об'єкт відноситься до даного класу, якщо він близький в плані евклідової відстані.

Класифікатори по мінімуму відстані з недіагональною матрицею коваріації.

В такому випадку максимізація $L_i(x) = -\frac{1}{2}(x - \mu_i)\Sigma_i^{-1}(x - \mu_i)^T$ забезпечує мінімізацію відстані Махаланобіса, яка визначається як:

$d_M = ((x - \mu_i)^T \Sigma^{-1}(x - \mu_i))^{1/2}$. Оскільки матриця коваріації є симетричною, то її доцільно представити в такому вигляді:

$\Sigma = \Phi \cdot \Lambda \cdot \Phi^T$, де $\Phi^T = \Phi^{-1}$, а Λ - діагональна матриця з власними значеннями матриці Σ на діагоналі. Матриця Φ має стовбці, які відповідають власним векторам матриці Σ : $\Phi = (v_1, \dots, v_k)$ на діагоналі.

Таким чином, отримуємо лінію рівновіддалених точок x :

$$(x - \mu_i)^T \cdot \Phi \cdot \Lambda^{-1} \cdot \Phi^T (x - \mu_i) = C^2$$

Нехай $x' = \Phi^T x$. Тоді координатами x' будуть $v_k^T x$, тобто проєкції вектора x на власні вектори. Іншими словами, ми отримали координати в новій системі координат, у якої осі визначаються власними векторами $v_k x$, $k = 1, 2 \dots l$. Тоді останнє рівняння перетвориться в рівняння еліпсоїда в новій системі координат:

$$\frac{(x'_1 - \mu'_{i1})^2}{\lambda_1} + \frac{(x'_2 - \mu'_{i2})^2}{\lambda_2} + \dots + \frac{(x'_l - \mu'_{il})^2}{\lambda_l} = C^2$$

При $l=2$, центр еліпса знаходиться в точці з координатами $\mu_i = (\mu_{i1}, \mu_{i2})$, а головні осі співпадають з власними векторами і мають довжини $2\sqrt{\lambda_1} \cdot C$ і $2\sqrt{\lambda_2} \cdot C$ відповідно.

Приклад №5:

Розглянемо двовимірний, двокласовий випадок класифікації двох нормально розподілених векторів з коваріаційною матрицею $\Sigma = \begin{pmatrix} 1.1 & 0.3 \\ 0.3 & 1.9 \end{pmatrix}$ і середніми значеннями $\mu_1 = (0, 0)^T$ і $\mu_2 = (3, 3)^T$.

Знайдемо Σ^{-1} :

$$|\Sigma| = 1.1 \cdot 1.9 - 0.3^2 = 2.09 - 0.09 = 2$$
$$\Sigma^{-1} = \frac{1}{2} \begin{pmatrix} 1.9 & -0.3 \\ -0.3 & 1.1 \end{pmatrix} = \begin{pmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{pmatrix}$$

Класифікуємо для прикладу вектор $(1.0; 2.2)$. Для цього визначимо відстань Махаланобіса:

$$\begin{aligned} d_m^2(\mu_1, x) &= (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) = \\ &= (1, 2.2) \begin{pmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{pmatrix} \begin{pmatrix} 1 \\ 2.2 \end{pmatrix} = \\ &= (0.95 - 0.33) + (-0.15 + 1.21) \cdot 2.2 = \\ &= 0.57 + 1 \cdot 0.6 \cdot 2.2 = 0.57 + 2.332 = 2.952 \\ d_m^2(\mu_2, x) &= (x - \mu_2)^T \Sigma^{-1} (x - \mu_2) = \\ &= (-1, -0.8) \begin{pmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{pmatrix} \begin{pmatrix} -1 \\ -0.8 \end{pmatrix} = \\ &= -(-1.9 + 0.12) - (0.3 - 0.44) \cdot 0.8 = \\ &= 3.56 + 0.112 = 3.672 \end{aligned}$$

Таким чином, хоча сама точка $(1.0; 2.2)$ по евклідовій відстані ближча до $(0, 0)$ ніж до точки $(3, 3)$, но по відстані Махаланобіса вона ближче до $(3, 3)$.

Тепер обчислимо головні осі еліпса з центром в точці $(0, 0)$. Для цього визначимо власні значення:

$$\begin{vmatrix} 1.1 - \lambda & 0.3 \\ 0.3 & 1.9 - \lambda \end{vmatrix} = 2.09 - 3\lambda + \lambda^2 - 0.09 = \lambda^2 - 3\lambda + 2 = 0$$
$$\lambda_1 = 1, \lambda_2 = 2.$$

Тоді власні вектори і напрямки головних осей еліпса матимуть вигляд:

$$V_1 = \left(\frac{3}{\sqrt{10}}, \frac{-1}{\sqrt{10}} \right)^T, V_2 = \left(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}} \right)^T.$$

Математична модель нейрона.

За основу моделі нейрона, взято принцип дії та влаштування біологічного нейрону:

(схема 1)



Схема 1. Структура біологічного нейрону.

Узагальнена схема нейрона представлена на рис.1.

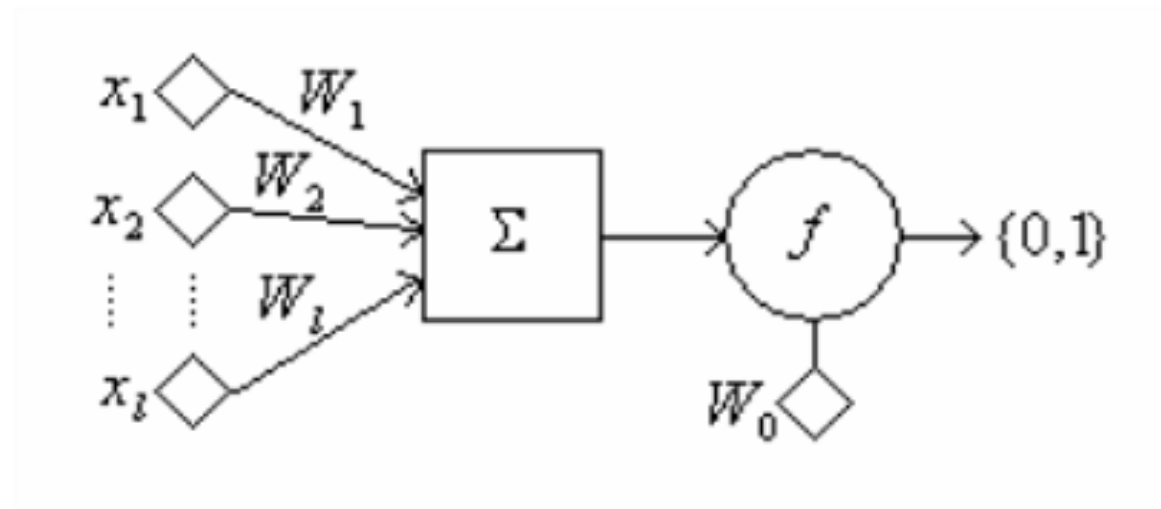



Рис.1. Узагальнена схема нейрона.

Тут x_1, x_2, \dots, x_n - компоненти вектора ознак $\vec{x} = (x_1, x_2, \dots, x_n)$, Σ - суматор, W_1, W_2, \dots, W_n - синаптичні ваги, f - функція активації, W_0 - поріг. Виходом суматора є

величина $\sum_{i=1}^l W_i \cdot x_i$, яка є входом (аргументом) функції активації. Значення функції

активації обчислюється на основі визначення знака суми $\sum_{i=1}^l W_i \cdot x_i + W_0$:

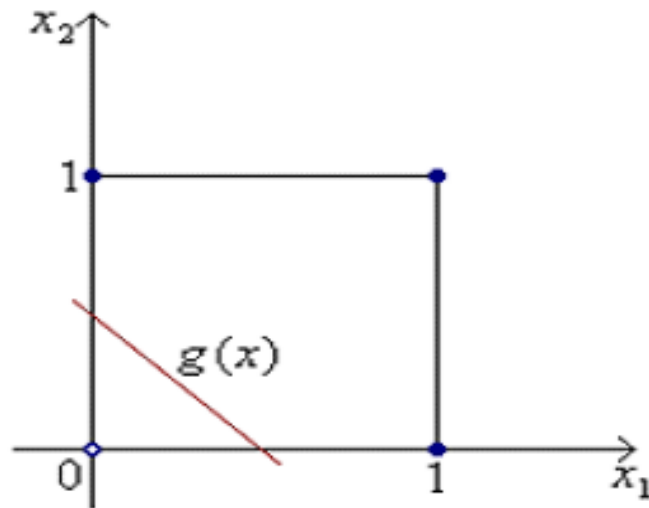
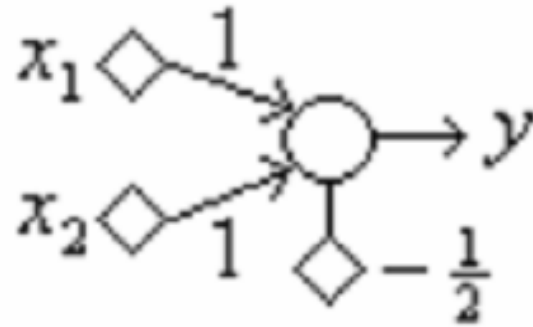
$$f(v) = \begin{cases} 0, \text{ при } v < 0; \\ 1, \text{ при } v > 0. \end{cases}$$



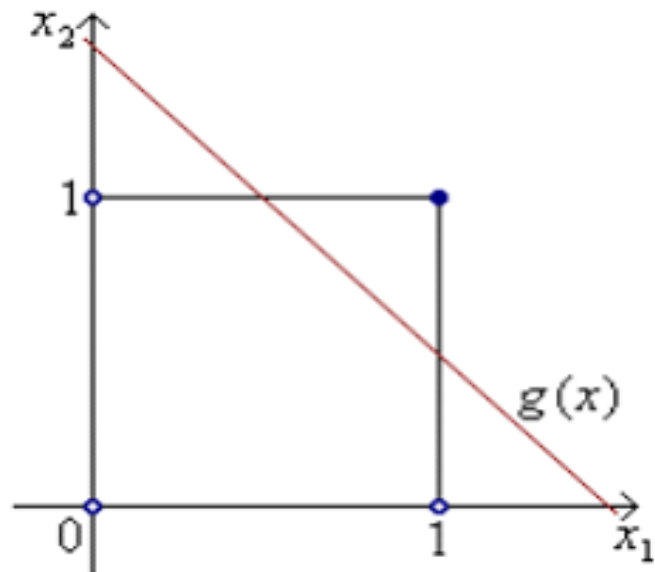
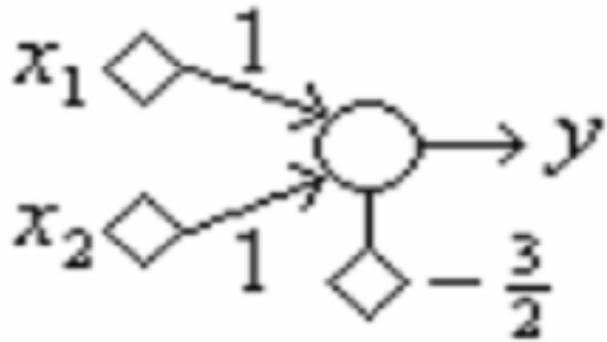
Таким чином, нейрон являє собою лінійний класифікатор з дискримінантною функцією $g(x) = \sum_{i=1}^l W_i \cdot x_i + W_0$.

Тоді задача побудови лінійного класифікатора для заданої множини прецедентів зводиться до задач навчання нейрона, тобто підбору відповідних ваг W_1, W_2, \dots, W_n і порогу W_0 . Навчання полягає в корекції синаптичних ваг і порогу.

1. Побудова лінійного класифікатора функції $og(x_1, x_2)$. Очевидно, що розподільчою лінією є $x_1 + x_2 = \frac{1}{2}$. Відповідний персеPTRон має вигляд:



Побудова лінійного класифікатора функції $\text{and}(x_1, x_2)$. Очевидно, що розподільчою лінією є $x_1 + x_2 = \frac{3}{2}$. Відповідний персептрон має вигляд:



Нагадаємо таблиці значень функцій $and(x_1, x_2)$ та $or(x_1, x_2)$:

№ прецедента	x_1	x_2	$and(x_1, x_2)$	$or(x_1, x_2)$
1	0	0	0	0
2	0	1	0	1
3	1	0	0	1
4	1	1	1	1

Задача XOR. Нелінійний класифікатор.

Побудова нелінійного класифікатора функції $\text{xor}(x_1, x_2)$. Нехай на виході персептрона для функції $\text{or}(x_1, x_2) - y_1$, а на виході персептрона $\text{and}(x_1, x_2) - y_2$.

Визначимо значення вектора (y_1, y_2) .

Вхідні вектора		OR	AND	XOR	Клас
x_1	x_2	y_1	y_2		
0	0	0	0	1	Ω_1
0	1	1	0	0	Ω_0
1	0	1	0	0	Ω_0
1	1	1	1	1	Ω_1

Позначивши класи як показано в таблиці, отримаємо розділяючу пряму, зображену на рис.2.

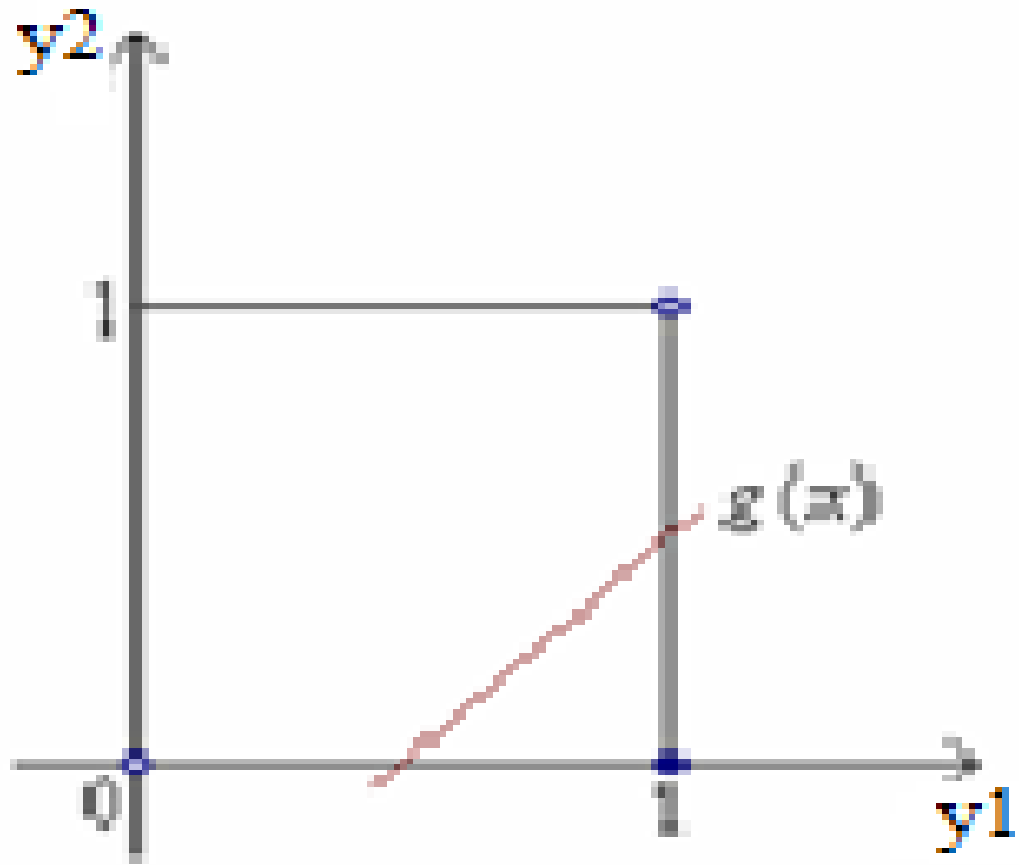


Рис.2. Розділяюча пряма для функції $\text{xor}(x_1, x_2)$.

Реалізація за допомогою $or(x_1, x_2)$ і $and(x_1, x_2)$, зображена на рис.3.

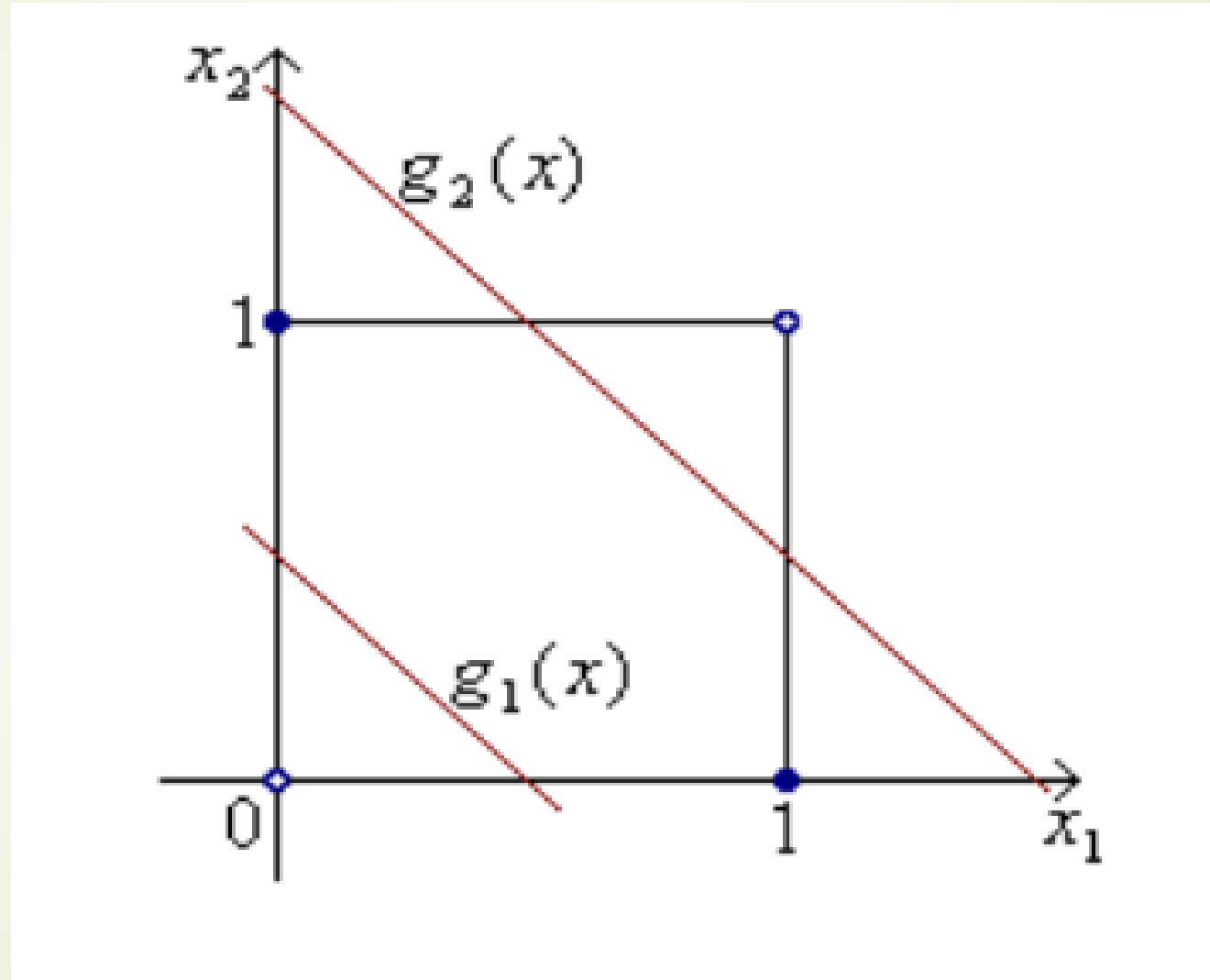


Рис.3 Реалізація за допомогою $or(x_1, x_2)$ і $and(x_1, x_2)$.