ЛЕКЦІЯ №5
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 n0 за несумісними гіпотезами H= (h1, h2,..., hλ,..., hΛ), де λ =1, 2,..., Λ – порядковий номер гіпотези, розділена на відповідні окремі групи nhλ, то з неї можна отримати формулу для повної групи ймовірностей, підсумок яких дорівнює 1.

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

де phλ = nhλ / n0 – ймовірність гіпотези hλ, λ =1, 2,..., Λ – порядковий номер гіпотези.

Якщо кожна група nhλ має деяку власну підгрупу nAλ з випадковими подіями Aλ, то отримуємо аналогічні вирази для повної групи умовних ймовірностей

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

де p(Aλ|A) = nAλ / nA= pAλ представляє умовну ймовірність: «ймовірність спостереження подій Aλ, якщо всі події A відбулись». Вертикальна риска у позначенні p(Aλ|A) відноситься до подій A і тлумачиться словом «якщо». З формул (5.4.2) отримуємо формули повної ймовірності:

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

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

Припустимо, що у вибірці спостережених вогнищ n0 ймовірності спостереження окремих подій А (полум'я), В (дим) та їх сумісного спостереження АВ (полум'я та дим) дорівнюють

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

де nA = nA1 + nAB – кількість спостережених подій А (полум'я); nB = nB1 + nAB – кількість спостережених подій В (дим); nAB – кількість одночасно спостережених подій (полум'я та дим), яка одночасно входить у множини nA та nB.

Умовні ймовірності спостереження події А (якщо події В вже відбулись) та події В (якщо події А вже відбулись) дорівнюють

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

Ймовірність одночасного спостереження двох подій дорівнює

$$p(AB) = \frac{n_{AB}}{n_0} = \frac{n_{AB}}{n_B} \cdot \frac{n_B}{n_0} = p(A \mid B)\, p(B); \tag{5.4.7}$$

$$p(AB) = \frac{n_{AB}}{n_0} = \frac{n_{AB}}{n_A} \cdot \frac{n_A}{n_0} = p(B \mid A)\, p(A). \qquad (5.4.8)$$

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

$$p(A \mid B) = \frac{p(B \mid A)\, p(A)}{p(B)}, \; p(B \mid A) = \frac{p(A \mid B)\, p(B)}{p(A)}. \qquad (5.4.9)$$

Напрямки аналізу ймовірнісних мереж за методом Байєса. Для розпізнавання кожної групи – гіпотези вектора H = (h1, h2,…, hλ,…, hΛ, де λ = 1, 2,…, Λ – порядковий номер гіпотези початкової вибірки об'єктів n0, використовують ознаки – змінні об'єктів вибірки X = (x1, x2,…, xj,…, xn), 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 x1;
– вибірка з кількома ознаками xj, j = 1,…, n.

Кожна з цих вибірок може мати задану кількість взаємно несумісних гіпотез H= (h1, h2,..., hλ,..., hΛ), де λ =1, 2,..., Λ – порядковий номер гіпотези, ймовірності яких phλ складають повну групу.

1. Вибірка з об'єктами, що описуються однією ознакою x1.

Одна ознака – змінна x1 може приймати одне з двох значень: або x11; hλ=1, або x10; hλ=0. Тому кожна група гіпотези розділяється на дві підгрупи у вигляді [(nh1; x11, nh1; x10),..., (nhλ; x11, nhλ; x10),..., (nhΛ; x11, nhΛ; x10)] з дотриманням рівності nhλ = nhλ; x11+ nhλ; x10.

Ми можемо:

- Визначити апріорну ймовірність спостереження гіпотези hλ, λ =1, 2,..., Λ за формулою:

$$p_{h\lambda,x11} = \frac{n_{h\lambda;x11}}{n_0} = \frac{n_{h\lambda}}{n_0} \cdot \frac{n_{h\lambda;x11}}{n_{h\lambda}} = p_{h\lambda} p(x_{11;h\lambda} \mid h_\lambda), \quad (5.4.11)$$

де $ph\lambda = nh\lambda / n0$ – апріорна ймовірність гіпотези $h\lambda$; $p(x11;h\lambda|h\lambda) = nh\lambda;x11/nh\lambda$ – умовна ймовірність спостереження ознаки $x11:h\lambda=1$ в групі $nh\lambda$ (гіпотези $h\lambda$); $p(x10;h\lambda|h\lambda) = 1 - p(x11;h\lambda|h\lambda)$ – умовна ймовірність спостереження ознаки $x10:h\lambda= 0$ (або $x10:h\lambda= -1$) в групі $nh\lambda$.

– Розрахувати ймовірність спостереження об'єктів з заданим значенням ознаки – змінної $x11;h\lambda= +1$ для всієї вибірки

$$p_{x11} = \sum_{\lambda=1}^{\Lambda} \frac{n_{h\lambda;x11}}{n_0} = \sum_{\lambda=1}^{\Lambda} \left(\frac{n_{h\lambda}}{n_0} \cdot \frac{n_{h\lambda;x11}}{n_{h\lambda}} \right) = \sum_{\lambda=1}^{\Lambda} p_{h\lambda} p(x_{11;h\lambda} \mid h_\lambda). \quad (5.4.12)$$

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 $x11:h\lambda=+1$ дорівнює

$$p(h_\lambda \mid x_{11}) = \frac{n_{h\lambda,x11}}{\sum_{\lambda=1}^{\Lambda} n_{h\lambda,x11}} = \frac{n_{h\lambda,x11}}{n_0 \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} n_{h\lambda,x11}} = \frac{p_{h\lambda,x11}}{p_{x11}} = \frac{p_{h\lambda} p(x_{11;h\lambda} \mid h_\lambda)}{\sum_{\lambda=1}^{\Lambda} p_{h\lambda} p(x_{11;h\lambda} \mid h_\lambda)}. \quad (5.4.13)$$

2. У вибірці з об'єктами, що описуються багатьма змінними X = (x1, x2,..., xj, ..., xn), при умові незалежності змінних, елементи вектора X для однієї гіпотези спостерігаються одночасно.

- Тому в формулі (5.4.13) ймовірність спостереження однієї гіпотези hλ 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 hλ при класифікації невідомого вектора X

$$p_{h\lambda}^{Д} == \frac{p_{h\lambda}\prod_{j=1}^{n} p(x_j \mid h_\lambda)}{\sum_{\lambda=1}^{\Lambda}\left[p_{h\lambda}\prod_{j=1}^{n} p(x_j \mid h_\lambda)\right]}, \qquad (5.4.14)$$

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

Тому що в навчальній групі $n_{h\lambda}$ гіпотези h_λ існують навчальні об'єкти з ознаками $x_j = +1$ та $x_j = -1$, то, внаслідок їх рівнозначущості 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 Х, 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(xj|hλ) дорівнює нулю, або близька до нуля, то порушується основний принцип статистики: одноосібне «рішення однієї змінної xj» у виразі $\prod_{j=1}^{n} p(x_j \mid h_\lambda)$ «забороняє» «рішення» всіх інших змінних, навіть якщо всі їх умовні ймовірності p(xj|hλ) дорівнюють 1 (тобто якщо всі інші змінні свідчать на користь прийняття гіпотези).

Цього недоліку позбавлена формули Байєсового класифікатора, які рекомендуються до використання

$$p_{h\lambda}^{\Sigma} = \frac{p_{h\lambda} \sum_{j=1}^{n} p(x_j \mid h_\lambda)}{\sum_{\lambda=1}^{\Lambda} \left[p_{h\lambda} \sum_{j=1}^{n} p(x_j \mid h_\lambda) \right]}, \qquad (5.4.15)$$

$$p_{h\lambda}^{p\Sigma} = p_{h\lambda} \sum_{j=1}^{n} p(x_j \mid h_\lambda), \qquad (5.4.16)$$

- 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 xj (ця незалежність математично означає неможливість розпізнати окремий об'єкт, який повинен мати індивідуальні залежні між собою ознаки); жодна змінна xj 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 актульним 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:

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

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.

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

де функція f – монотонно зростає.

Визначення.

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

Таким чином, поверхня рішень, буде задаватися рівнянням:

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

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 $N(\mu_i, \Sigma_i)$:

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

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 – рядкі.

Квадратична поверхня рішень.

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

$$g_i(x) = \ln(P(\Omega_i|x)) =$$

$$= \ln(p(x|\Omega_i)P(\Omega_i)) =$$

$$= \ln p(x|\Omega_i) + \ln P(\Omega_i) =$$

$$= -\frac{1}{2}\frac{(x-\mu_i)}{\Sigma_i}(x-\mu_i)^{\mathrm{T}} + \ln P(\Omega_i) + \ln\frac{1}{(2\pi)^{l/2}|\Sigma_i|^{1/2}} =$$

$$= -\frac{1}{2}\frac{(x-\mu_i)}{\Sigma_i}(x-\mu_i)^{\mathrm{T}} + \ln P(\Omega_i) - \frac{l}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i| =$$

$$= -\frac{1}{2}\frac{(x-\mu_i)}{\Sigma_i}(x-\mu_i)^{\mathrm{T}} + \ln P(\Omega_i) + C_i, \text{ где } C_i = -\frac{l}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i|$$

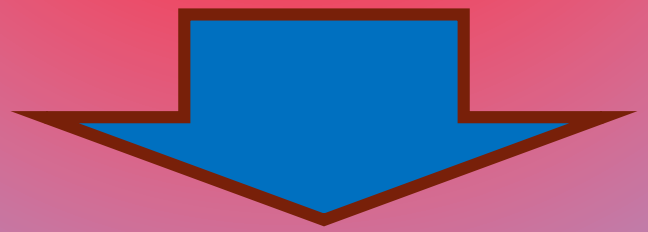
Ця функція, являє собою квадратичну форму. Значить розділяюча поверхня $g_i(x) - g_j(x) = 0$ являє собою гіперповерхню другого порядку. Тому Байєсовський класифікатор є квадратичним класифікатором.

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

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.

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

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

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

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

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

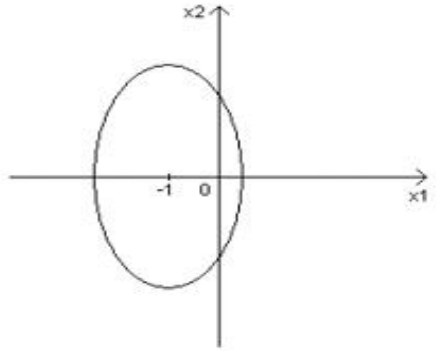
$$= -\frac{1}{2}\left(5(x_1 + 1)^2 + \frac{8}{3}x_2^2\right) + 5 + \frac{1}{2}\ln\frac{10}{3}$$

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

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

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

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



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

Звернути увагу на матриці коваріації ознак обох класів

Приклад №3:

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

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:

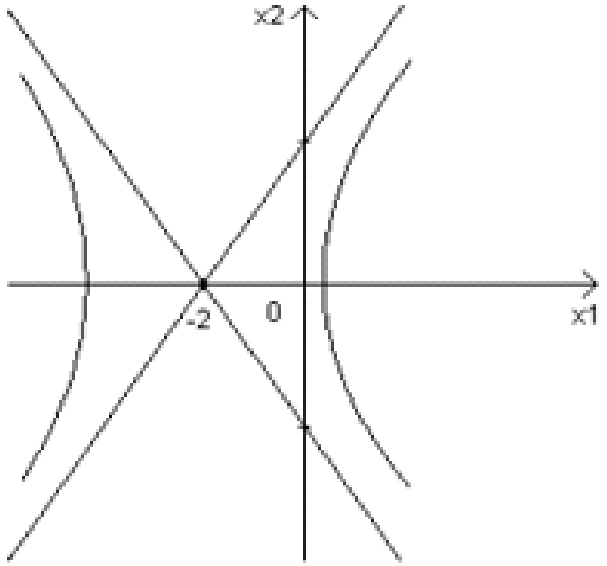
$$g_1(x) = -\frac{1}{2}\left(5(x_1 - 1)^2 + 4x_2^2\right) + \ln P(\Omega_2) - \ln(2\pi) + \frac{1}{2}\ln 20$$

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

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

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

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



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

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

Лінійна поверхня рішення.

Для тієї ж умови:

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

Вище, ми отримали квадратичну форму:

$$g_i(x) = \ln(P(\Omega_i|x)) =$$

$$= \ln(p(x|\Omega_i)P(\Omega_i)) =$$

$$= \ln p(x|\Omega_i) + \ln P(\Omega_i) =$$

$$= -\frac{1}{2} \frac{(x-\mu_i)}{\Sigma_i} (x-\mu_i)^T + \ln P(\Omega_i) + \ln \frac{1}{(2\pi)^{l/2} |\Sigma_i|^{1/2}} =$$

$$= -\frac{1}{2} \frac{(x-\mu_i)}{\Sigma_i} (x-\mu_i)^T + \ln P(\Omega_i) - \frac{l}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma_i| =$$

$$= -\frac{1}{2} \frac{(x-\mu_i)}{\Sigma_i} (x-\mu_i)^T + \ln P(\Omega_i) + C_i, \text{ где } C_i = -\frac{l}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma_i|$$

Нехай $\Sigma_i = \Sigma_j$, тоді:

$$h_i(x) = -\frac{1}{2}\left[\frac{x}{\Sigma_i}x^{\mathrm{T}} - \frac{\mu_i}{\Sigma_i}x^{\mathrm{T}} - \frac{x}{\Sigma_i}\mu_i^{\mathrm{T}} + \frac{\mu_i}{\Sigma_i}\mu_i^{\mathrm{T}}\right] + \ln P(\Omega_i) + C_i =$$

$$= -\frac{1}{2}\left[\frac{x}{\Sigma_i}x^{\mathrm{T}} - 2\frac{\mu_i}{\Sigma_i}x^{\mathrm{T}} + \frac{\mu_i}{\Sigma_i}\mu_i^{\mathrm{T}}\right] + \ln P(\Omega_i) + C_i =$$

$$= -\frac{1}{2}\left[K_i(x) - 2W_i x^{\mathrm{T}} + W_i\mu_i^{\mathrm{T}}\right] + \ln P(\Omega_i) + C_i =$$

$$= -\frac{1}{2}K_i(x) + L_i(x) + C_i, \text{ где } L_i(x) = W_i x^{\mathrm{T}} + W_{i0};\ W_i = \frac{\mu_i}{\Sigma_i};$$

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

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:

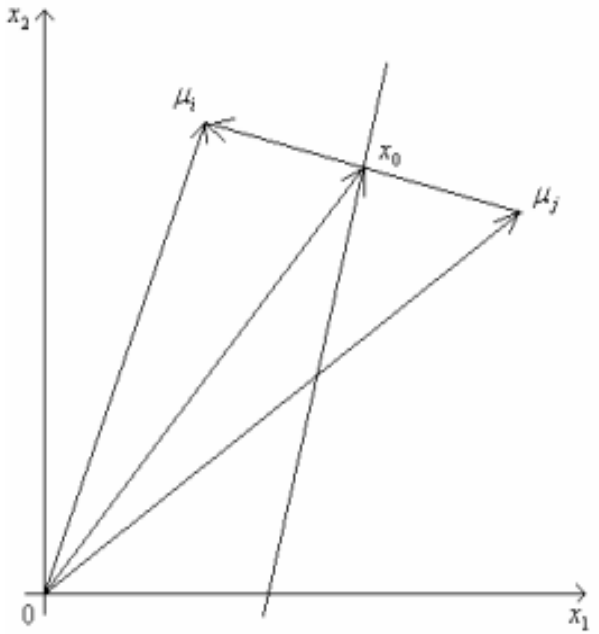
$\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. Тоді $L_i(x)$ матиме вигляд:

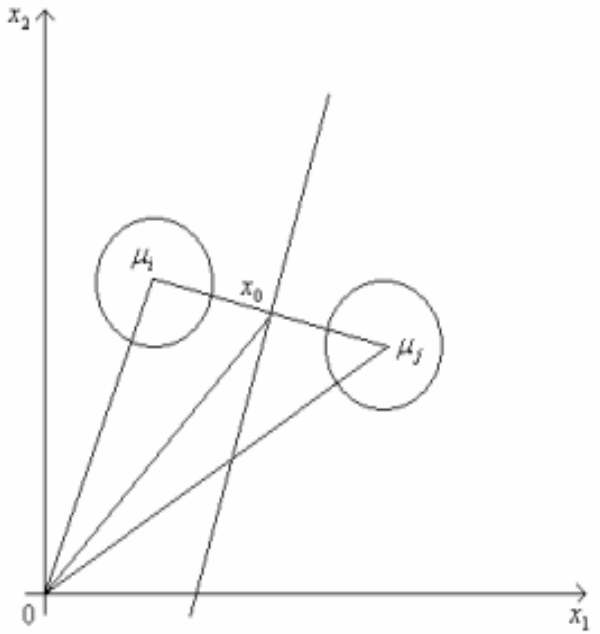
$$L_i(x) = \frac{1}{\sigma^2}\mu_i^T x + W_{i0};\ L_{ij}(x) = L_i(x) - L_j(x) = W^T(x - x_0) = 0,$$

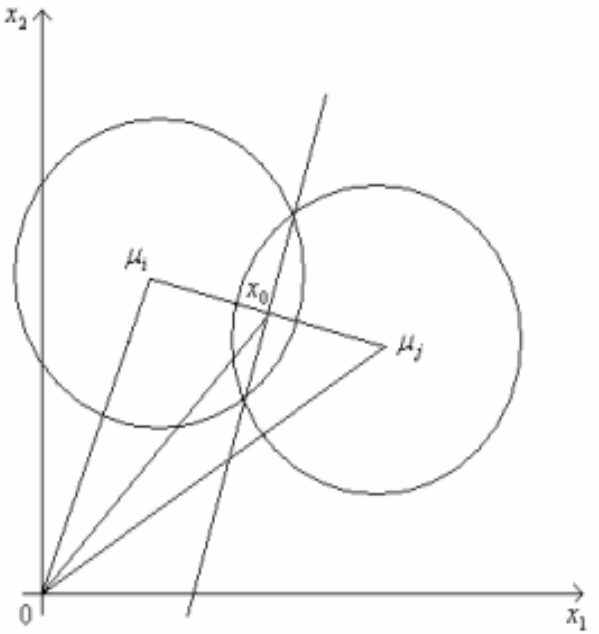
де $W = \mu_i - \mu_j$, $x_0 = \frac{1}{2}(\mu_i + \mu_j) - \sigma^2 \frac{\mu_i - \mu_j}{\|\mu_i - \mu_j\|^2} \ln\left(\frac{P(\Omega_i)}{P(\Omega_j)}\right)$. В даному випадку під нормою розуміється евклідова норма. Поверхнею рішення, є **гіперплощина**, яка проходить через точку x_0.

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 оргогональна вектору $\overline{\mu_i \mu_j}$. Схеми нижче демонструють різні положення лінійної розділяючої поверхні.

РОЗТАШУВАННЯ ДИСКРИМІНАНТНОЇ ФУНКЦІЇ В ГРОСТОРІ ОЗНАК ЧІТКО СПОСТЕРІГАЄТЬСЯ ВПЛИВ ДИСПЕРСІЇ КЛАСОВИХ ОЗНАК







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

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

буде мати дещо інші параметри $W = \frac{\mu_i - \mu_j}{\Sigma};\ \ x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\mu_i - \mu_j}{\|\mu_i - \mu_j\|^2_{\Sigma^{-1}}}$. 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 $\Sigma_1 = \Sigma_2 = ... = \Sigma_n = \Sigma$ 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:

$d_M = \left((x-\mu_i)^T \Sigma^{-1}(x-\mu_i)\right)^{1/2}$. 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: $\Phi = (\nu_1, \ldots \nu_k)$ на діагоналі.

Таким чином, отримаємо 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 $\nu_k^T x$, тобто проекції вектора x на власні вектори. Іншими словами, ми отримали координати в новій системі координат, у якої осі визначаються власними векторами $\nu_k x$, $k = 1,2..l$. Тоді останнє рівняння перетвориться в рівняння еліпсоїда в новій системі координат:

$$\frac{(x_1' - \mu_{i1}')^2}{\lambda_1} + \frac{(x_2' - \mu_{i2}')^2}{\lambda_2} + \ldots + \frac{(x_l' - \mu_{il}')^2}{\lambda_l} = C^2$$

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:

$$d_m^2(\mu_1, x) = (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) =$$

$$= (1,2.2)\begin{pmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{pmatrix}\begin{pmatrix} 1 \\ 2.2 \end{pmatrix} =$$

$$= (0.95 - 0.33) + (-0.15 + 1.21) \cdot 2.2 =$$

$$= 0.57 + 1 \cdot 0.6 \cdot 2.2 = 0.57 + 2.332 = 2.952$$

$$d_m^2(\mu_2, x) = (x - \mu_2)^T \Sigma^{-1} (x - \mu_2) =$$

$$= (-1,-0.8)\begin{pmatrix} 0.95 & -0.15 \\ -0.15 & 0.55 \end{pmatrix}\begin{pmatrix} -2 \\ -0.8 \end{pmatrix} =$$

$$= -(-1.9 + 0.12) - (0.3 - 0.44) \cdot 0.8 =$$

$$= 3.56 + 0.112 = 3.672$$

Таким чином, хотя 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 напрямок головних осей еліпса матимуть вигляд:

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

ШТУЧНИЙ НЕЙРОН.

За основу моделі нейрона, взято принцип дії та влашлування біологічного нейрону. (схема 1)

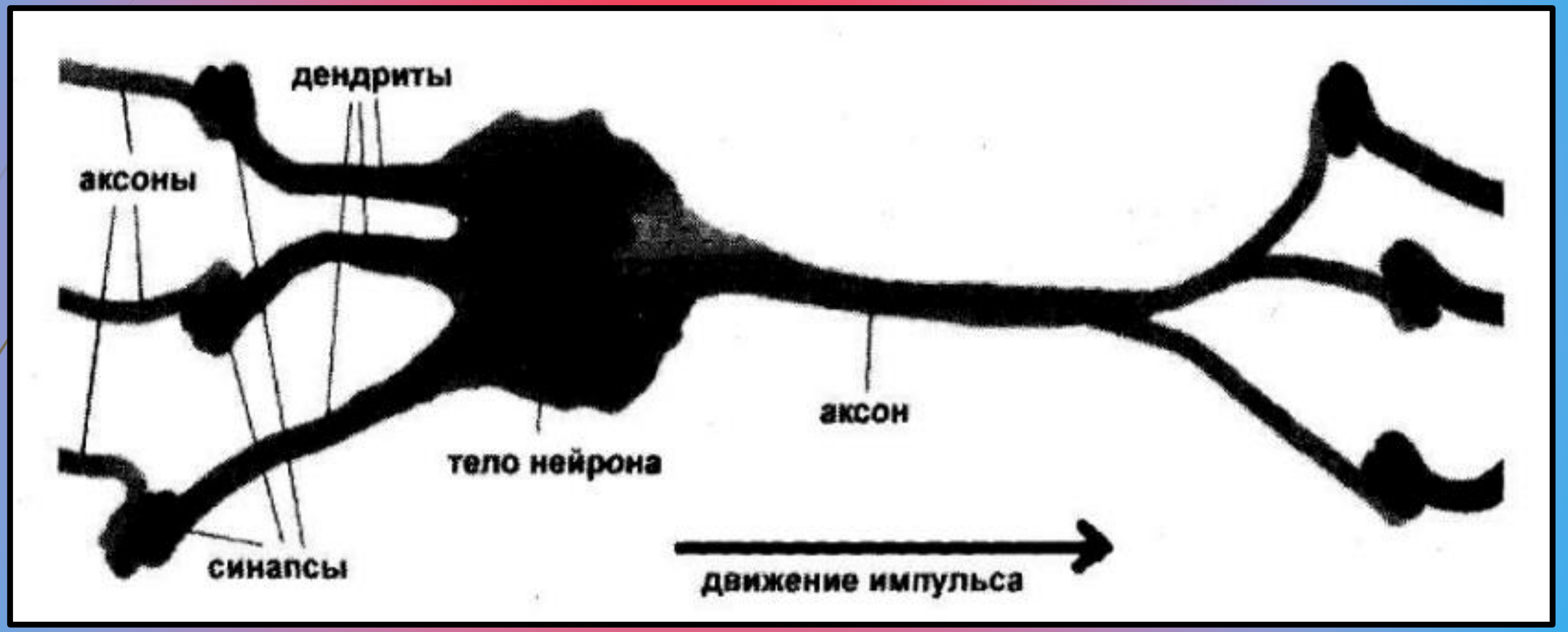


Схема 1. Структура біологічного нейрону.

Узагальнена схема нейрона представлена на рис.1.

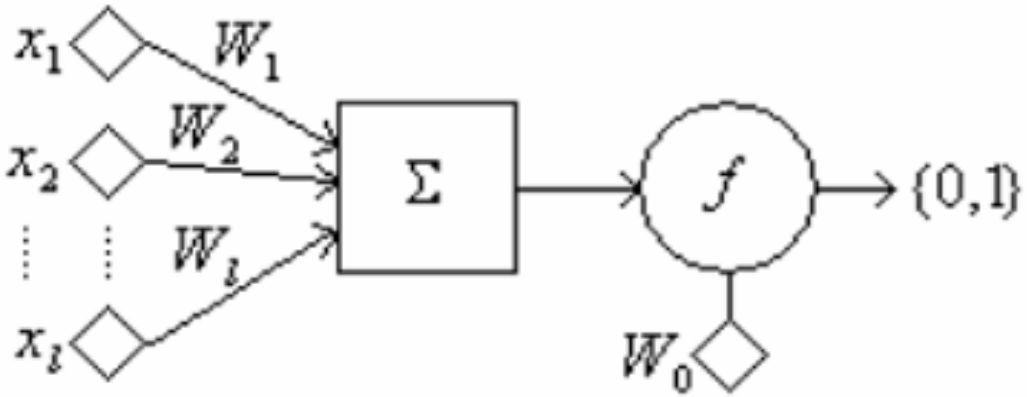


Рис.1. Узагальнена схема нейрона.

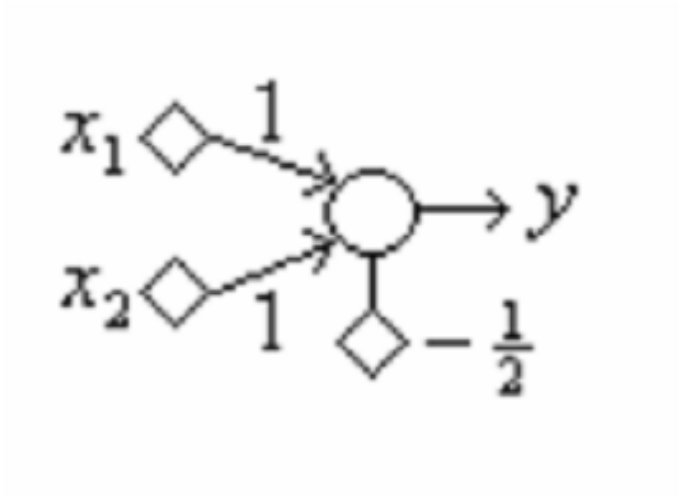
Тут $x_1, x_2, \ldots\ldots x_n$ - компоненти вектора ознак $\vec{x} = (x_1, x_2, \ldots\ldots x_n)$, Σ - суматор, $W_1, W_2, \ldots\ldots W_n$ - синаптичні ваги, f - функція активації, W_0 - поріг. Виходом суматора є величина $\sum_{i=1}^{l} W_i \cdot x_i$, яка є входом (аргументом) функції активації. Значення функції активації обчислюється на основі визначення знака суми $\sum_{i=1}^{l} W_i \cdot x_i + W_0$:

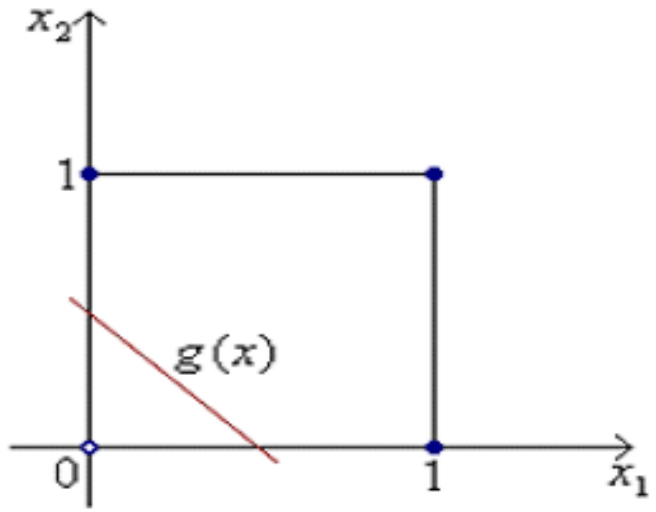
$$f(\nu) = \begin{cases} 0, \textit{при} \quad \nu < 0; \\ 1, \textit{при} \quad \nu > 0. \end{cases}$$

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

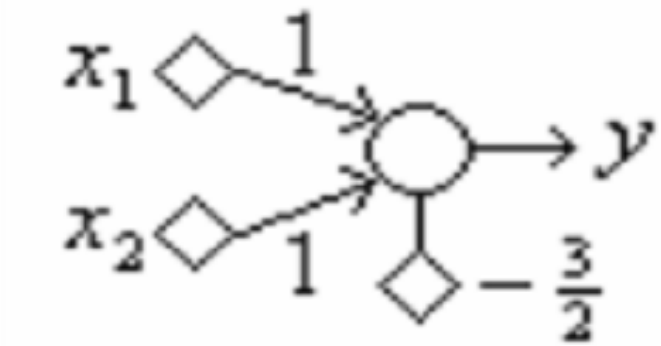
Тоді задача побудови лінійного класифікатора для заданої множини прецендентів зводиться до задач навчання нейрона, тобто підбору відповідних ваг $W_1, W_2, \ldots . W_n$ і порогу W_0. Навчання полягає в корекції синаптичних ваг і порогу.

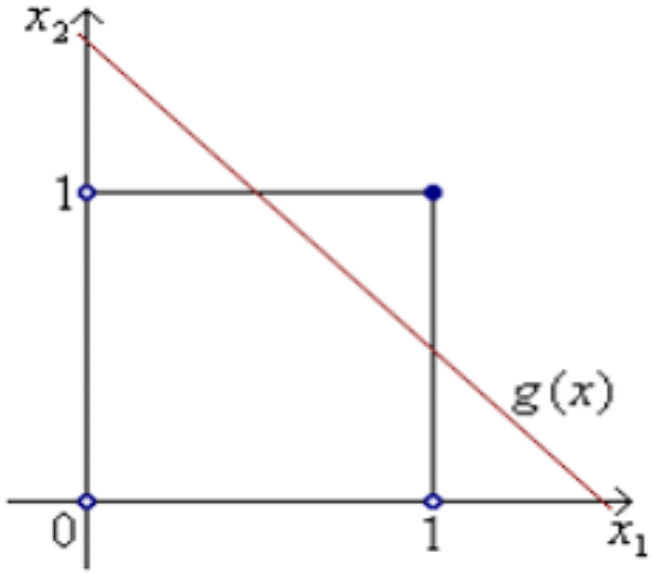
1. Побудова лінійного класифікатора функції or(x1, x2). Очевидно, що розподільчою лінією є $x_1 + x_2 = \frac{1}{2}$. Відповідний персептрон має вигляд:





Побудова лінійного класифікатора функції and(x1, x2). Очевидно, що розподільчою лінією є $x_1 + x_2 = \frac{3}{2}$. Відповідний персептрон має вигляд:





Нагадаємо таблиці значень функцій and(x1, x2) та or(x1, x2):

№ 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 xor(x1, x2). Нехай на виході персептрона для функції or(x1, x2) – y1, а на виході персептрона and(x1, x2) – y2.

Визначимо значення вектора (y1, y2).

Вхідні вектора		OR	AND	XOR	
x_1	x_2	y_1	y_2		Клас
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.

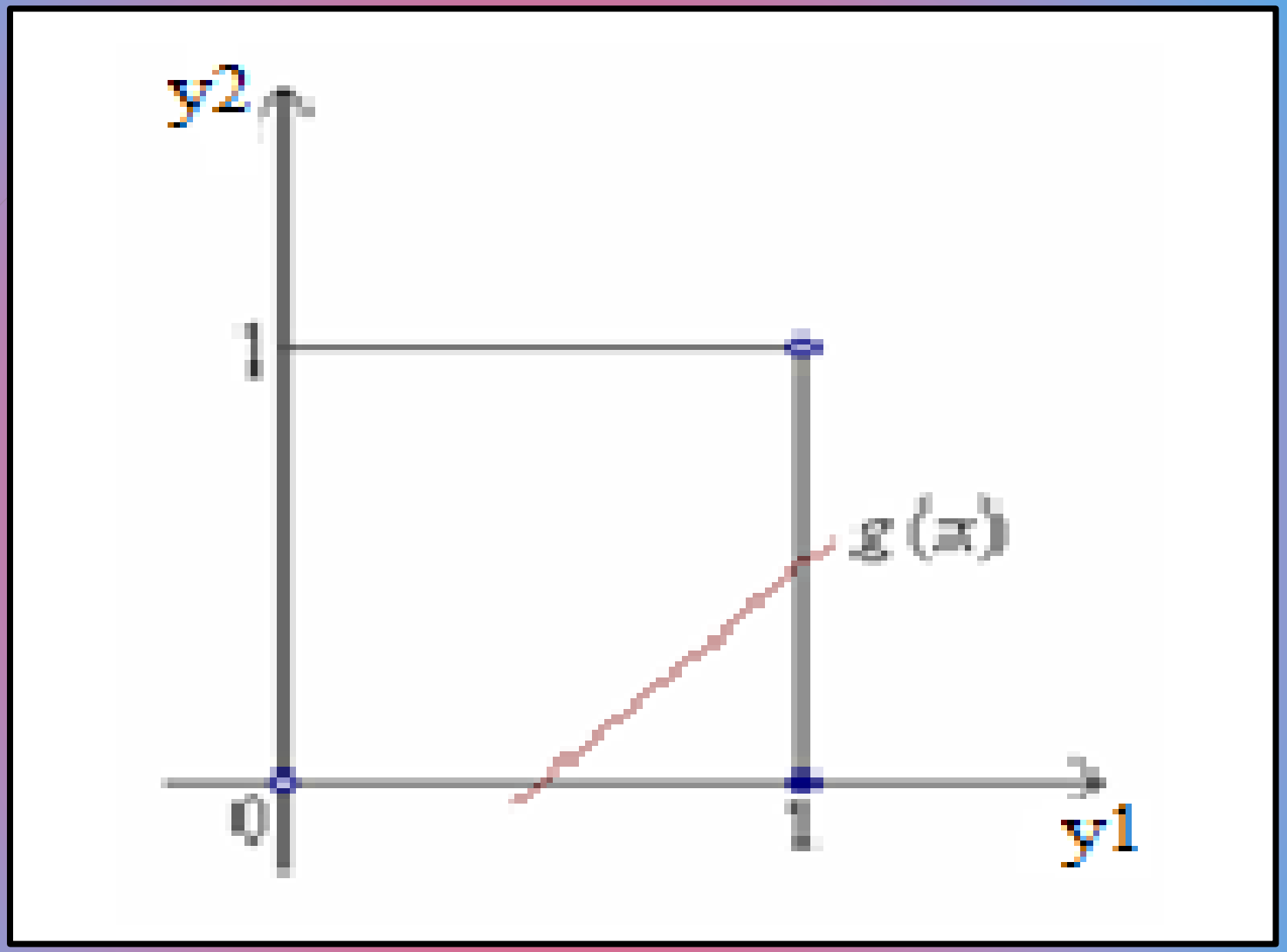


Рис.2. Розділяюча пряма для функції xor(x1, x2).

Реалізація за допомогою or(x1, x2) і and(x1, x2), зображена на рис.3.

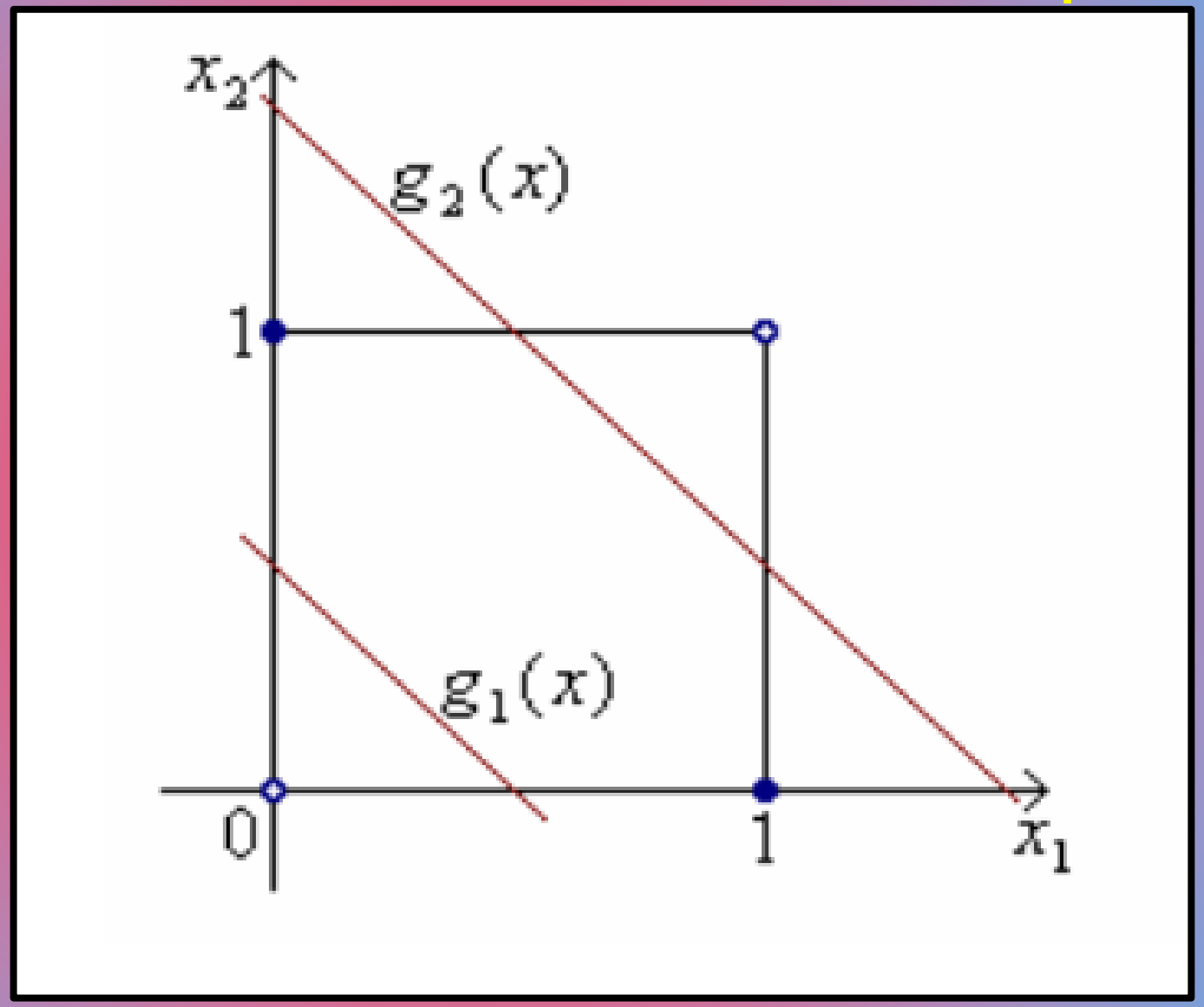


Рис.3 Реалізація за допомогою or(x1, x2) і and(x1, x2).