# Лабораторна робота №9

**Тема:** «Нейронні мережі. Особливості вирішальних функцій».

**Мета:** Дослідження класифікаційних особливостей нейронних мереж. Вирішальні функції.

**Час виконання:** 2 години.

**Навчальні питання:**

1). Класифікаційні особливості двухслойного персептрона;

2). Особливості вирішальних функцій.

## Теоретична частина

**Класифікаційні особливості двухслойного персептрона.**

Розглянемо загальний випадок двухслойного персептрона (рис.1).

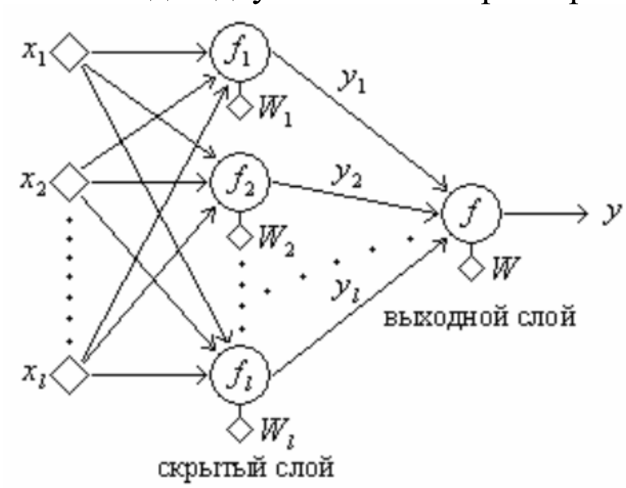


Рис.1. Двухшаровий персептрон.

Нехай $x \in R^l$ і в прихованому слої $p$ нейронів. Прихований слой нейронів відображає $R^l$ в $H_p \in R^p$, де $H_p = \{(y_1, y_2, .....y_p) \in R^p, y_i \in [0, 1], 1 \le i \le p\}$ - гіперкуб. Іншими словами, кожний нейрон задає гіперплощину, яка розділяє простір навпіл, тобто прихований слой нейронів ділить простір $R^l$ на поліедри. Всі вектори з кожного поліедра відображаються в вершину $p$ – мірного одиничного куба. Вихідний нейрон розділяє вектори в класах описаних поліедрами, тобто виконує переріз гіперкуба, отриманого в прихованому слої.

**Приклад 1.**

Розглянемо нейронну сіть з двома входами ($l = 2$) і трьома нейронами ($k = 3$). Тоді простір $R^l = R^2$. Нехай перший слой нейронів задає розбиття ознакового простору (площина) як показно на рис. 2.

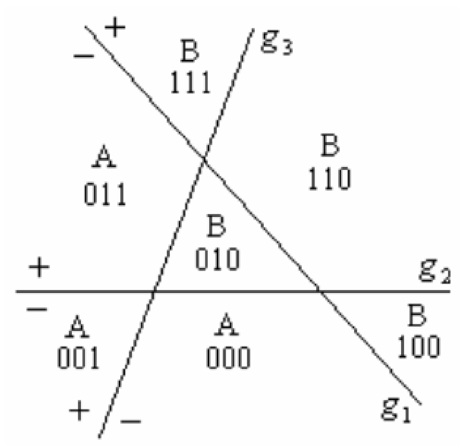


Рис.2. Розбиття ознакового простору (площина).

В кожному багатокутнику (може бути і нескінченний) всі точки відповідають одному класу (А або В). При цьому, в кожному багатокутнику знаки лінійних функціоналів $g_1$, $g_2$, $g_3$ зостаються постійними.

Значить, з кожним багатокутником зв'язано встановлене значення вектора виходів нейронів першого шару, причому для різних багатокутників ці значення різні. Оскільки значеннями компонент цього вектора є 0 або 1, отримаємо, що кожному багатокутнику відповідає деяка вершина одиничного куба $H^3$ в просторі $R^3$. При цьому, кожній вершині куба відповідає один клас А або В.

На рис.3 зображено одиничний куб $H^3$, в якого темні вершини відносяться до класу А, а світлі вершини до класу В. Задача нейронів другого шару полягає в розділенні вершин цього куба.

В нашому прикладі, площина $y_1 + y_2 - y_3 = \dfrac{1}{2}$ є розділяючою для куба $H^3$. Ця площина і задає параметри нейрона другого шару. Відмітимо, що вершина (1, 0, 1) в кубі не загружена, тобто в неї не відображається ні один багатокутник.

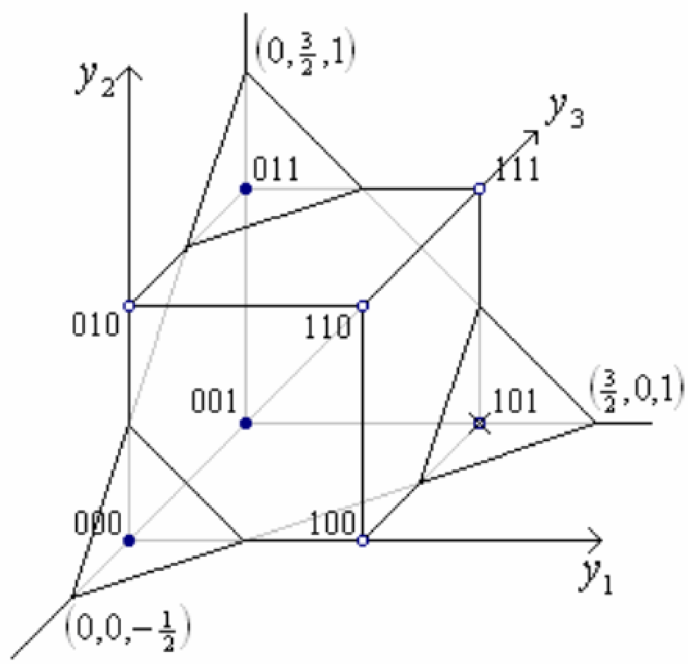


Рис.3. Поділ одиничного куба $H^3$ розділяючою площиною.

**Трьохслойний персептрон.**

Зовнішній (вихідний) нейрон реалізує лише одну гіперплощину.Одна розділяюча гіперплощина не завжди забезпечує бажаний розділ вершин гіперкуба. Наприклад, якщо два кінця його однієї з двох головних діагоналей відносяться до класу А, а два кінця другої діагоналі до класу В. З такою ситуацією ми стикалися в задачі XOR. Введемо ще один шар нейронів (рис.4)

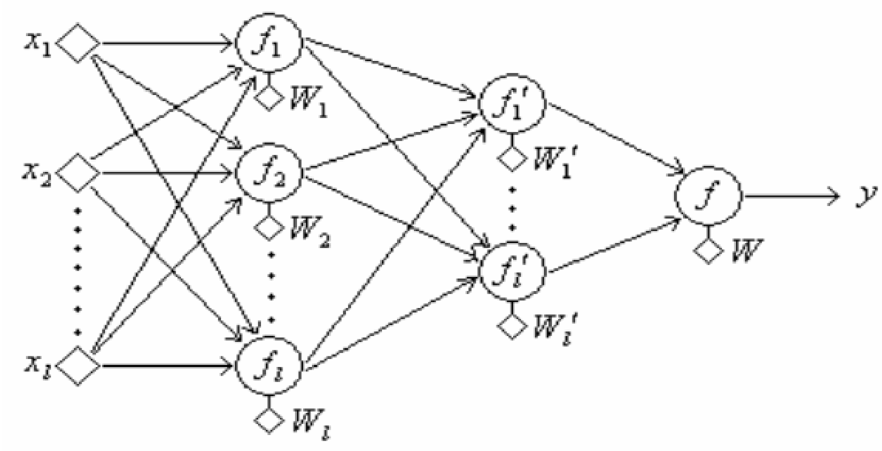


Рис.4. Трьохшаровий персептрон.

**Визначення.**

Трьохшарова нейронна мережа дозволяє описати довільні розділи об'єднань поліедрів.

**Доведення:**

Розглянемо перший шар з $p$ нейронів. На першому шарі формуються гіперплощини, оскільки формується поліедральне розбиття простору гіперплощинами. Для заданої скінченної множини прецендентів завжди можна побудувати розбиття простору ознак на поліедри таке, що **ні в якому поліедрі не знайдеться пари точок з різних класів.** Як показано вище, перший шар відображає поліедри в вершини $p$ – мірного одиничного гіперкуба.

**Оскільки з кожним поліедром пов'язані образи одного класу, то і з кожною вершиною гіперкуба пов'язаний один клас.**

Кожний нейрон другого шару виконує переріз отриманого гіперкуба. Оберемо в якості таких перерізів гіперплощини, які відсікають рівно одну вершину гіперкуба. Оскільки кількість вершин гіперкуба дорівнює $2^p$ то кількість нейронів другого шару також дорівнює $2^p$. Таким чином, вихід **нейронів** другого шару представляє собою вектор розмірності $2^p$, у якого завжди одне значення дорівнює 1, а інші дорівнюють нулю. Назвемо нейрони другого шару нейронами класу А або В у відповідності з класом вершини гіперкуба, яку відсікає цей нейрон. Тепер очевидно, яким чином формувати третій шар нейронної мережі. Необхідно, у вихідному нейроні третього шару реалізувати **оператор логічного складання виходів нейронів другого шару, які відносяться до класу А.** таким чином, розділяюча гіперплощина вихідного нейрона задається рівнянням:

$$c_1 z_1 + c_2 z_2 + ....... + c_k z_k = \frac{1}{2}, \quad \text{де } k = 2^p, \text{ а } c_i = \begin{cases} 1 \text{ якщо нейрон відноситься до класу } A \\ 0 \text{ в іншому випадку} \end{cases}$$

Таким чином, можна побудувати трьохшаровий персептрон наступним чином. Нейрони першого шару розділяють простір ознак на поліедри одного класу і відображають їх у вершини гіперкуба. Нейрони другого шару відсікають вершини гіперкуба. Нейрон третього шару виконує класифікацію через оператор логічного складання. Таким чином визначення доведено.

Розглянемо як будується рівняння гіперплощини, що відсікає вершину $p$ – мірного одиничного гіперкуба. Діагональ гіперкуба має довжину $\sqrt{p}$. Довжини діагоналей *(p-1)* – мірних одиничних гіперкубів, які є бічними гранями $p$ – мірного одиничного гіперкуба, дорівнюють $\sqrt{p-1}$. Центр куба знаходиться в точці $\left(\frac{1}{2},\frac{1}{2},.......\frac{1}{2}\right)$. Відстань від центра куба до довільної вершини дорівнює $\frac{\sqrt{p}}{2}$. Площину проводимо перпендикулярно головній діагоналі куба, інцидентну вершині, яку потрібно відсікти, так щоб відстань від цієї вершини до січної площини становило $\frac{\sqrt{p}-\sqrt{p-1}}{2}$, причому дана точка повинна знаходитись на діагоналі куба, проведеної до відсікаємої вершини.

**Побудова нейронної мережі.**

Існує два підходи до задачі побудови нейронної мережі класифікатора. Перший підхід полягає в побудові мережі змінюючи архітектуру. Даний метод базується на точній класифікації прецендентів. Другий підхід полягає в підборі параметрів (ваги і пороги) для мережі заданої архітектури.

**Алгоритми, які базуються на точній класифікації множин прецендентів.**

За основу береться **один** нейрон. Далі нарощуємо нейрони, поки не отримаємо правильну класифікацію всіх прецендентів (рис.5).

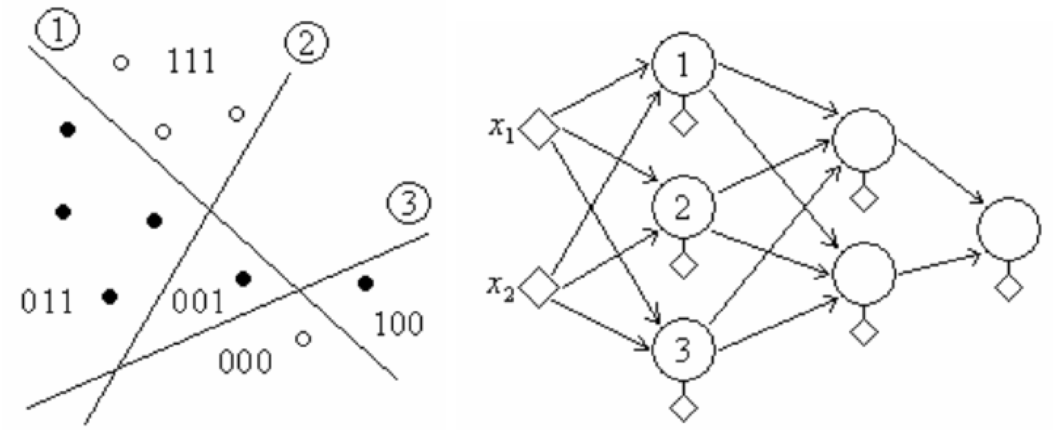


Рис.5 класифікація прецендентів.

Починаємо з одного нейрона *n(X)*, який називається **майстром**. Після його тренування отримуємо розподіл множини $X$ на $X^+$ і $X^-$. Якщо $X^+$ включає вектори з двох класів, то вводимо новий вузол $n(X^+)$, який називаються **послідовником**.

Таким чином, на першому шарі знаходиться один майстер і декілька послідовників. **Ніякі вектори з різних класів не мають однакового виходу з першого шару.**

$X_1 = \{y : y = f_1(x), x \in X\}$, де $f_1$ - відображення яке задає перший шар. Аналогічно будується другий шар, третій шар, і.т.д.

**Визначення.**

При правильному виборі ваг, кожен черговий шар правильно класифікує всі вектори, які правильно класифікував майстер і ще хотя би один нейрон.

Таким чином, отримаємо архітектуру, яка має скінченне число шарів, що правильно класифікують всі преценденти (рис.6).

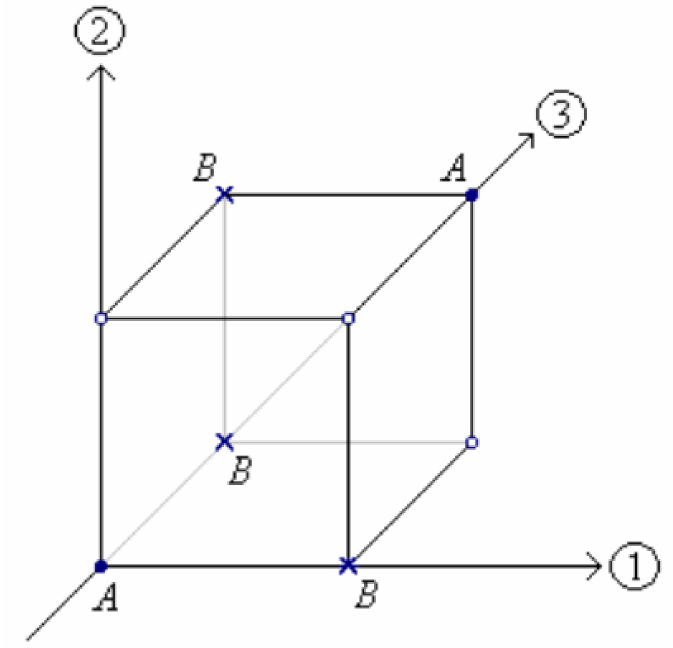


Рис.6. Класифікація преценлентів.

**Особливості вирішальних функцій.**

***Вирішальні функції***. Припустимо, що нам відомі області існування трьох класів $\Omega_1$, $\Omega_2$, $\Omega_3$. Звичайно при проектуванні НМ кількість класів, що розпізнається збільшується на одиницю: додатковий клас $\Omega_4$ має назву «Не розпізнаний об'єкт» і призначений для об'єктів, які не розпізнаються. Для пояснення такого принципу розпізнавання розглянемо класи $\Omega_1$, $\Omega_2$, $\Omega_3$ на площині ознак $\{x_1, x_2\}$ (рис. 6.1.1).

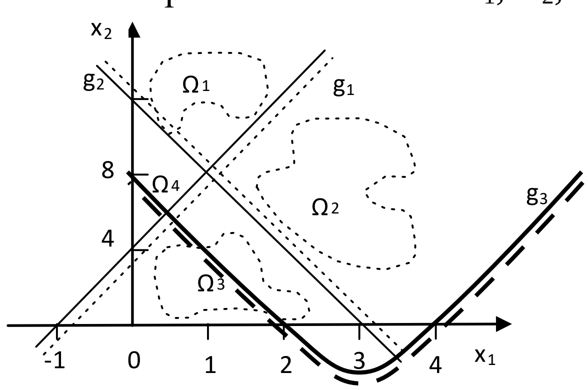


Рис. 6.1.1. Простір ознак класів $\Omega_1 – \Omega_4$

Вказані на рис. 6.1.1 ділянки класів можуть бути розділені двома довільними безперервними прямими лініями $g_1$ та $g_2$, які і є вирішальними (розділюючими) функціями чотирох класів (з врахуванням додаткового класу $\Omega_4$ «Не розпізнаний об'єкт»). Після накреслення цих двох довільних прямих ліній, що роздяляють класи, визначають вирішальні (розділюючі) функції цих прямих ліній по відрізках, які вони відсікають на осях координат.

Для функції $g_1$, отримаємо рівності $\frac{x_1}{-1}+\frac{x_2}{4}=1;\ 4x_1-x_2=-4;\ g_1^0=4x_1-x_2+4=0$, останню з яких перетворюємо у нерівність (за традицією знак рівності «=» замінюють на знак нерівності «≥»; але з рівним успіхом можна використовувати й знак «≤»)

$$g_1^0=4x_1-x_2+4\geq 0 \qquad (6.1.1)$$

Аналогічно для функції $g_2$ отримуємо $\frac{x_1}{3}+\frac{x_2}{12}=1;\ 4x_1+x_2=12;$ звідки отримаємо:

$$g_2=4x_1+x_2-12\geq 0 \qquad (6.1.2)$$

Кожна з двох прямих (6.1.1) та (6.1.2) розділяє площину ознак $\{x_1, x_2\}$ на дві напівплощини: в одній напівплощині нерівність виконується, а в іншій – ні.

Наприклад, для точки $\{x_1= x_2=0\}$ згідно (6.1.1) величина $g_1=4>0$. Ця нерівність виконується для всій напівплощини, у якій знаходиться точка $\{x_1= x_2=0\}$. У цих твердженнях можна пересвідчитись розрахунками для будь-яких інших довільних точок всієї площини ознак $\{x_1, x_2\}$. Напівплощина, у якій виконується нерівність $g_1>0$, позначається пунктирною лінією. Тоді відповідно для іншої напівплощини, не позначеною пунктирною лінією, нерівність $g_1>0$ не виконується. Таким чином, достатньо отримати значення функції для будь-якої довільної точки, щоб визначити, у якій напівплощині виконується нерівність $g_1>0$. Звичайно така перевірка виконується для точки $\{x_1= x_2=0\}$.

Таким же самим чином позначається пунктирною лінією напівплощина, у якій виконується нерівність $g_2>0$.

***Вирішальні правила*** використовують вирішальні (розділяючі) функції для визначення класів (див. рис. 6.1.1):

1. Клас $\Omega_1$: якщо $g_1<0$ та $g_2\geq 0$;
2. Клас $\Omega_2$: якщо $g_1\geq 0$ та $g_2\geq 0$;
3 Клас $\Omega_3$: якщо $g_1\geq 0$ та $g_2<0$; (6.1.3)
4. Клас $\Omega_4$: якщо $g_1<0$ та $g_2<0$.

У цьому випадку для визначення класів ми використали нерівності, які, як відомо з математики, виконують логічні функції. Можливо отримати й апаратну реалізацію виразів (6.1.1) – (6.1.3).

У принципі вирішальні функції можуть бути не лише прямими лініями, але й *нелінійними*, наприклад, мати вигляд *параболи* $g_3=x_1^2-x_2-6x_1+8\geq 0$ (рис. 6.1.1), при цьому пунктирна парабола частку півплощини, в якій нерівність виконується.

***Максимізація мінімальної відстані вирішальної функції від двох сусідніх класів***. По принципу дії апарат вирішальних функцій та вирішальних правил мало відрізняється від нейронних мереж (НМ), які виконують аналогічний логічний аналіз по розділу класів з аналогічним математичним та графічним поясненнями. Для обох напрямків аналізу ***розміщення вирішальних функцій є невизначеним***: з рис. 6.1.1 видно, що для розділу класів можна використати безліч вирішальних функцій. Виникає проблема визначення «оптимальних» вирішальних функцій. Метод вирішальних правил розглядається окремо від НМ, бо він у порівнянні з середовищем НМ з її більш жорсткими обмеженнями дозволяє зручніше виділити (шляхом математичних перетворень) «*оптимальні*» вирішальні функції. З цією метою використовується, наприклад, *метод «опорних векторів»*, за яким у багатомірному гіперпросторі у двох сусідніх класах виділяють найближчі сусідні «опорні вектори» двох різних класів, між якими й визначають положення розділюючої їх «оптимальної» гіперплощини (цей метод використовується також й для розділу багатьох класів).

**Вирішальні правила для складного розміщення класів.**

Вирішальні функції та вирішальні правила використовуються у n-вимірному просторі кількісних ознак *{$x_1$, $x_2$,…,$x_n$}*. Але далеко не завжди у таких НМ класи розміщуються таким чином, щоб між ними було зручно визначати вирішальні функції. Припустимо, що нам відомі області існування трьох класів $\Omega_1$, $\Omega_2$, $\Omega_3$ на площині ознак*{$x_1$, $x_2$}* (рис. 6.2.1), але класи $\Omega_1$, $\Omega_2$, $\Omega_3$ «переплутані» між собою.

У цьому випадку ми можемо довільними вирішальними функціями «$g_1$ >0», «$g_2$ >0», «$g_3$ >0» розділити між собою класи. На базі отриманих вирішальних функцій та відомого розміщення класів згідно рис. 6.2.1 отримуємо вирішальні правила для визначення класів:

1. ***Клас $\Omega_1$***, якщо (($g_1$ >0 та $g_3$ <0) або ($g_1$ <0 та $g_3$ <0)) або ((($g_1$ <0 та $g_2$ <0) та $g_3$ >0) або ($g_1$ <0 та $g_2$ >0)).
2. ***Клас $\Omega_2$***, якщо ($g_1$ >0 та $g_2$ >0).
3. ***Клас $\Omega_3$***, якщо (($g_1$ >0 та $g_2$ <0) та $g_3$ >0).

Можна впевнитись, що вирішальні функції однозначно визначають потрібні класи.

Таким чином, для класифікації об’єктів не є завадою те, що вирішальна функція пересікає область існування класу. У зв’язку з цим всю область існування реальних ознак всіх класів можна розділити на однакові комірки, для кожної з яких можна визначити клас за вирішальними правилами.

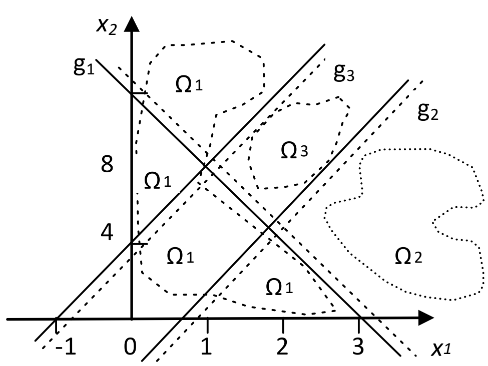


Рис. 6.2.1. Простір ознак класів $\Omega_1$, $\Omega_2$, $\Omega_3$.

З рис. 6.2.1 видно, що ***розміщення вирішальних функцій є невизначеним***: для розділу класів можна використати безліч вирішальних функцій.

**Завдання:**

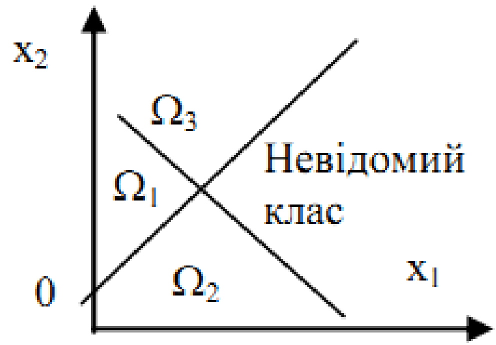


Рис.6.2.2. Розміщення трьох класів.

У площині 1-го квадранту з осей $x_1$, $x_2$ (рис. 6.2.2), на яких позначаються виміряні ознаки при максимальних значеннях $x_{1max} = N$, $x_{2max} = N$ де $N$ – порядковий номер студента у групі, виділити двома довільними прямими лініями три довільні ділянки (класи)*:* $\Omega_1$, $\Omega_2$, $\Omega_3$.

Четверту ділянку розділу вважати невідомим класом.

Отримати:

− вирішальні функції та вирішальні правила для визначення класів;

− схему апаратної реалізації системи РО;

− налагоджену програму (на будь-якій мові) із використанням вирішальних функцій та вирішальних правил.

*Примітка. Для отримання вирішальної функції використовують рівняння* $x_1/a_1 + x_2/b_2 = 1$ *(тут* $a_1$, $b_2$ *– координати точок пересічення прямою вирішальної функції відповідно осей* $x_1$, $x_2$*) з наступним переведенням його у нерівність вигляду* $u = b_2x_1 + a_1x_2 \geq a_1 b_2$.

У налагодженій програмі використати генератор випадкових чисел для генерації ознак $x_1$,$x_2$ об’єктів у визначених межах.

Захист лабораторної роботи передбачає виконання поставлених завдань у повному обсязі.