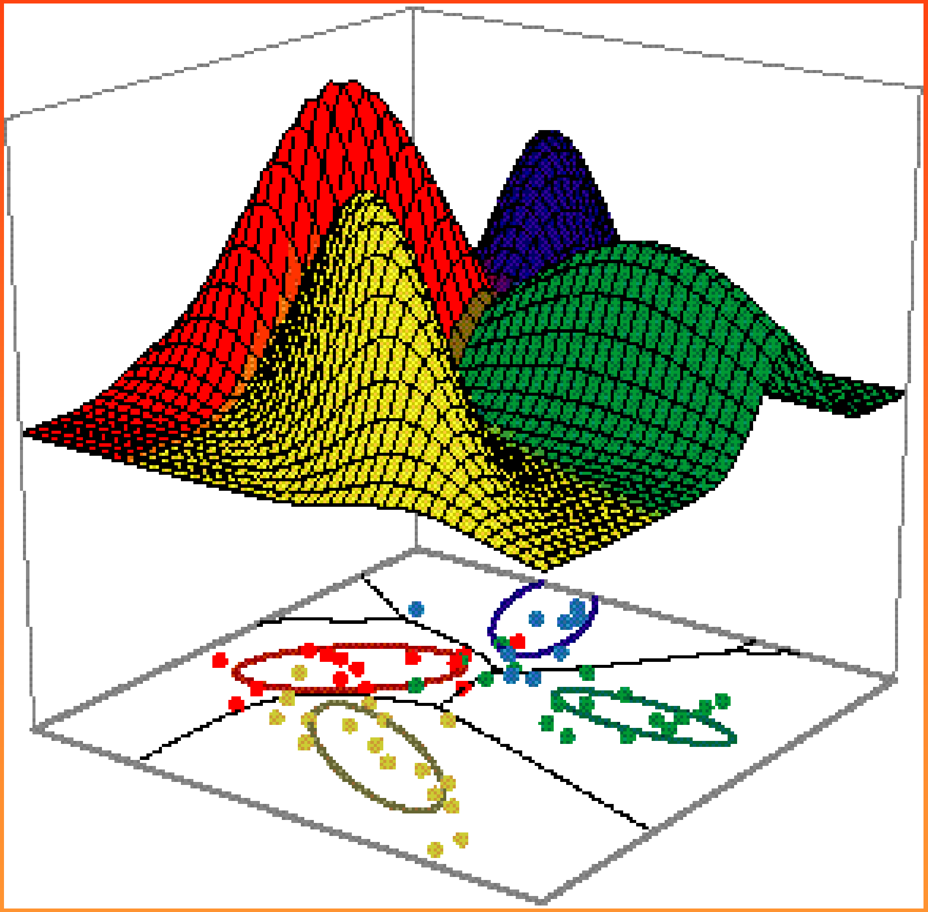

Pattern Classification

All materials in these slides were taken from *Pattern Classification (2nd ed) by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000*
with the permission of the authors and the publisher

Chapter 5:
Liniowe funkcje dyskryminujące
(Sections 5.1-5-3)

- Wprowadzenie
- Liniowe funkcje dyskryminujące i powierzchnie decyzyjne
- Funkcje dyskryminujące liniowe ze względu na parametry

Wprowadzenie

- Poprzednio zakładaliśmy znajomość rozkładów prawdopodobieństwa
- Obserwacje z ciągu uczącego były użyte do estymacji parametrów tych rozkładów (metoda największej wiarygodności, maksimum a'posteriori)
- Teraz zakładamy znajomość jedynie postaci funkcji dyskryminujących, podobnie jak w technikach nieparametrycznych
- Liniowe funkcje dyskryminujące są bardzo wygodne do obliczeń

Liniowe funkcje dyskryminujące i powierzchnie decyzyjne

- Definicja

  Liniowa funkcja decyzyjna to taka, która jest liniową kombinacją o postaci:

  $$g(x) = w^T x + w_0 \qquad (1)$$

  gdzie w jest wektorem współczynników wagowych a w_0 nazywamy progiem

- Klasyfikator binarny (rozpoznający dwie klasy) z funkcją dyskryminującą o postaci (1) korzysta z reguły decyzyjnej:

  Wybierz ω_1 jeżeli $g(x) > 0$ i ω_2 jeżeli $g(x) < 0$

  $\Leftrightarrow$ Wybierz ω_1 jeżeli $w^T x > -w_0$ i ω_2 w przeciwnym przyp.

  Jeżeli $g(x) = 0 \Rightarrow$ wynik rozpoznawania jest nieokreślony

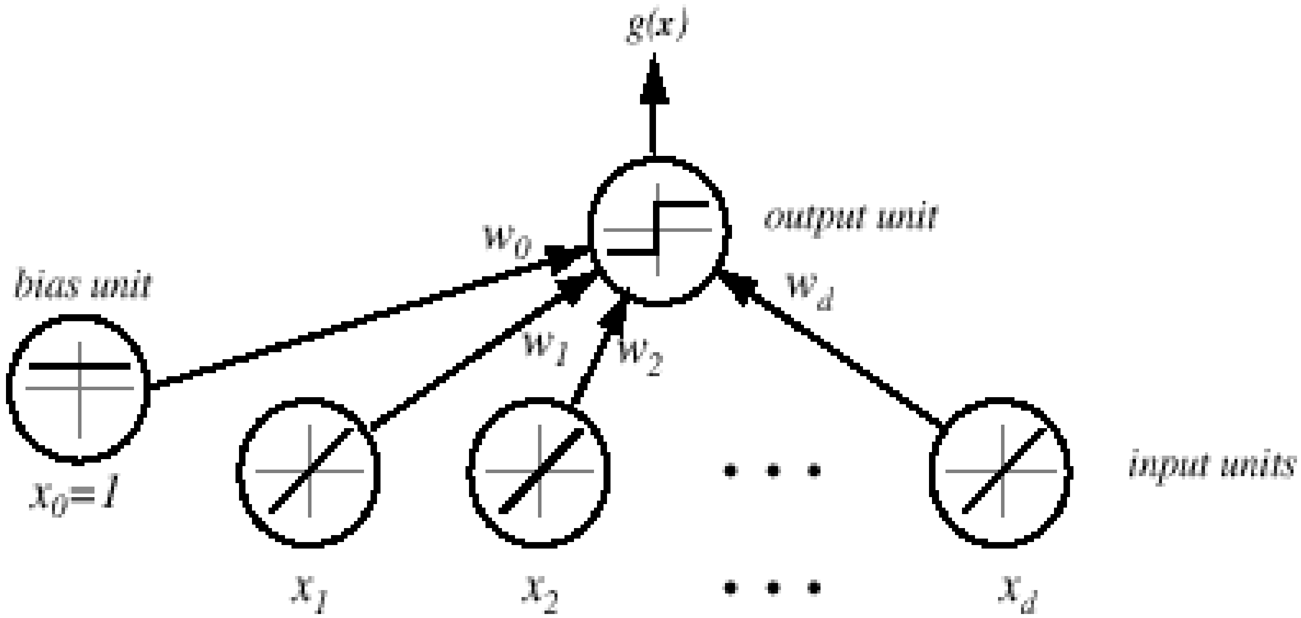

FIGURE 5.1. A simple linear classifier having d input units, each corresponding to the values of the components of an input vector. Each input feature value x_i is multiplied by its corresponding weight w_i; the effective input at the output unit is the sum all these products, $\sum w_i x_i$. We show in each unit its effective input-output function. Thus each of the d input units is linear, emitting exactly the value of its corresponding feature value. The single bias unit unit always emits the constant value 1.0. The single output unit emits a +1 if $\mathbf{w}^t\mathbf{x} + w_0 > 0$ or a −1 otherwise. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*.

- Równanie *g(x) = 0* definiuje powierzchnię decyzyjną separującą punkty należące do klasy ω_1 od punktów należących do klasy ω_2

- Dla liniowej funkcji *g(x)* powierzchnia decyzyjna to hiperpowierzchnia

- Funkcja dyskryminująca *g(x)* daje miarę odległości x od hiperpowierzchni

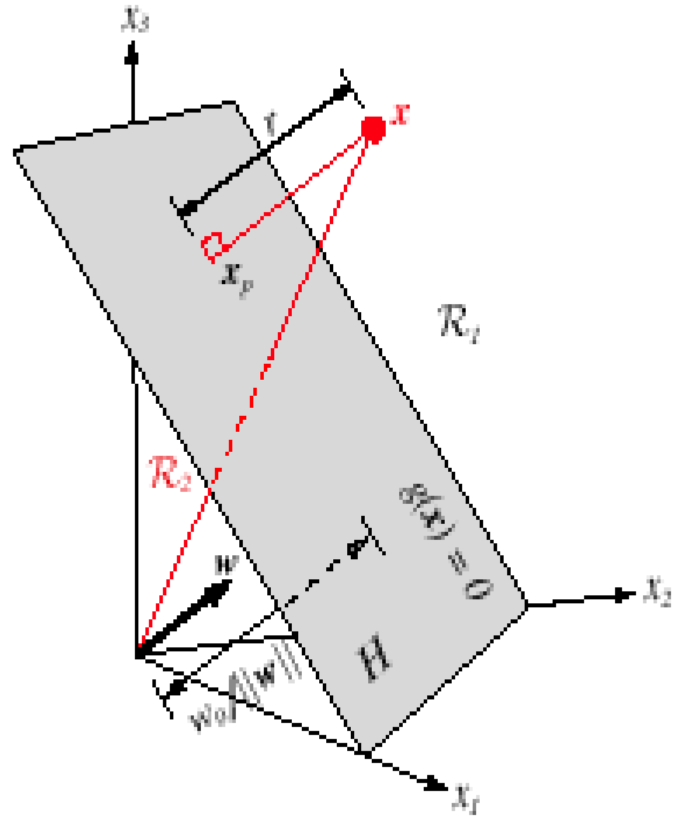

FIGURE 5.2. The linear decision boundary H, where $g(\mathbf{x}) = \mathbf{w}^t\mathbf{x}+w_0 = 0$, separates the feature space into two half-spaces $\mathcal{R}_1$ (where $g(\mathbf{x}) > 0$) and $\mathcal{R}_2$ (where $g(\mathbf{x}) < 0$). From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

$$x = x_p + \frac{r\mathbf{w}}{\|\mathbf{w}\|}$$ (ponieważ wektor $\mathbf{w}$ jest współliniowy

z wektorem $\mathbf{x - x}_p$ oraz $\frac{\mathbf{w}}{\|\mathbf{w}\|} = 1$

Ponieważ $g(\mathbf{x}) = 0$ i $\mathbf{w}^T\mathbf{w} = \|\mathbf{w}\|^2$ to $r = \frac{g(\mathbf{x})}{\|\mathbf{w}\|}$

W szczególności $d(\mathbf{0}, H) = \frac{\mathbf{w}_0}{\|\mathbf{w}\|}$

- Liniowa funkcja dyskryminująca dzieli przestrzeń cech hiperpowierzchnią decyzyjną
- Orientację tej powierzchni wyznacza wektor normalny *w* a jej położenie zależy od progu

Przypadek wielu klas

- Definiujemy c liniowych funkcji dyskryminujących

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + \mathbf{w}_{i0} \qquad i = 1, \ldots, c$$

  i przypisujemy x do klasy ω_i jeżeli $g_i(x) > g_j(x) \ \forall j \neq i$;
- Taki klasyfikator nazywamy „maszyną liniową”
- Maszyna liniowa dzieli przestrzeń cech na c obszarów decyzyjnych, przy czym $g_i(x)$ przyjmuje wartość największą dla x należącego do obszaru R_i
- Dla dwóch przyległych obszarów R_i , R_j, granica między nimi jest fragmentem hiperpłaszczyzny H_{ij} zdefiniowanej jako:

$$g_i(x) = g_j(x)$$

$$\Leftrightarrow \quad (w_i - w_j)^T x + (w_{i0} - w_{j0}) = 0$$

- $w_i - w_j$ jest normalny do H_{ij} i

$$d(\mathbf{x}, H_{ij}) = \frac{g_i - g_j}{\left\| \mathbf{w}_i - \mathbf{w}_j \right\|}$$

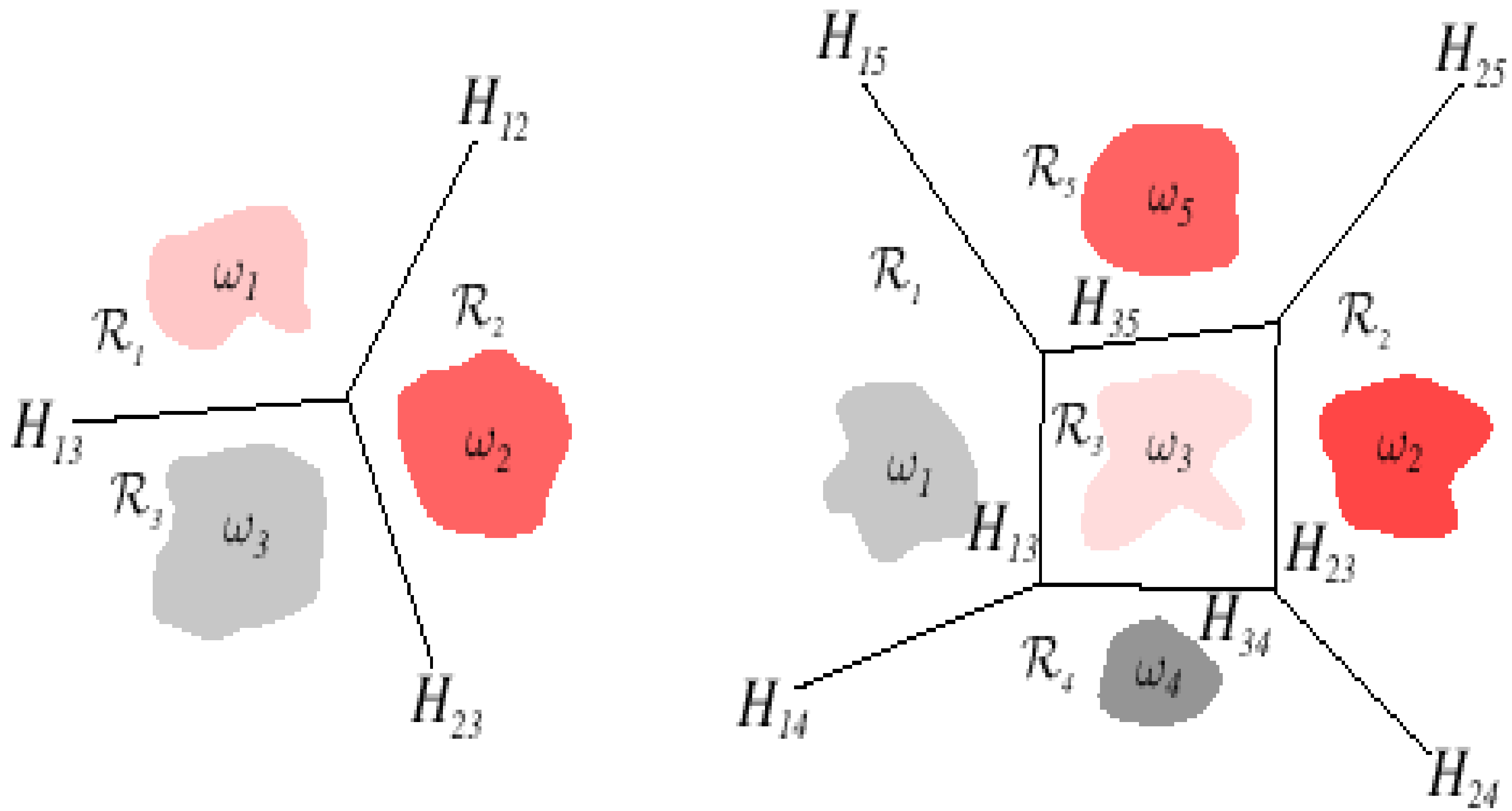

FIGURE 5.4. Decision boundaries produced by a linear machine for a three-class problem and a five-class problem. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- Można łatwo pokazać, że obszary decyzyjne generowane przez maszynę liniową są wypukłe. Nakłada to pewne ograniczenie na elastyczność i dokładność klasyfikatora

Funkcje dyskryminujące liniowe ze względu na parametry

- Granice obszarów decyzyjnych na ogół nie są liniowe
- Złożony kształt granic obszarów decyzyjnych wymaga użycia silnie nieliniowych funkcji
- Najczęściej rozważana jest ogólniejsza postać liniowych funkcji dyskryminujących:

$$g(x) = w_1 f_1(x) + w_2 f_2(x) + \ldots + w_N f_N(x) + w_{N+1} \qquad (1)$$

gdzie $f_i(x)$, $1 \leq i \leq N$ są jednowymiarowymi funkcjami wektora x,

$x \in R^n$ *(przestrzeń euklidesowa)*

- Wstawiając $f_{n+1}(x) = 1$, otrzymujemy:

$$g(x) = \sum_{i=1}^{N+1} w_i f_i(\mathbf{x}) = \mathbf{w}^T \dot{\mathbf{x}}$$

$$\text{gdzie } \mathbf{w} = (w_1, w_2, \ldots, w_N, w_{N+1})^T \text{ i } \dot{\mathbf{x}} = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_N(\mathbf{x}), f_{N+1}(\mathbf{x}))^T$$

- Ten zapis funkcji $g(x)$ wskazuje na fakt, że każda funkcja dyskryminująca postaci (1) może być traktowana jako liniowa w $(N + 1)$ – wymiarowej przestrzeni $(N + 1 > n)$

- $g(x)$ maintains its non-linearity characteristics in R^n

- Częstym wyborem funkcji *g(x)*, dla których $f_i(x)$ $(1 \leq i \leq N)$ są wielomiany

$$g(\mathbf{x}) = (\dot{\mathbf{w}})^T \dot{\mathbf{x}}$$

gdzie $\dot{\mathbf{w}}$ jest nowym wektorem parametrów wagowych wyznaczonym na podstawie oryginalnego wektora w oraz $f_i(x)$, $1 \leq i \leq N$

- Kwadratowe funkcje dyskryminujące dla 2-wymiarowej przestrzeni cech:

$$g(\mathbf{x}) = w_1 x_1^2 + w_2 x_1 x_2 + w_3 x_2^2 + w_4 x_1 + w_5 x_2 + w_6,$$
$$\mathbf{w} = (\mathrm{w}_1, w_2, ..., w_6)^T \text{ i } \dot{\mathbf{x}} = (\mathrm{x}_1^2, x_1 x_2, x_2^2, x_1, x_2, 1)^T$$

- Dla obrazów $x \in R^n$, najogólniejsza postać kwadratowych funkcji dyskryminujących jest następująca:

$$g(\mathbf{x}) = \sum_{i=1}^{n} w_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} x_i x_j + \sum_{i=1}^{n} w_i x_i + w_{n+1} \qquad (2)$$

Liczba elementów po prawej stronie wyrażenia jest równa:

$$l = N + 1 = n + \frac{n(n-1)}{2} + n + 1 = \frac{(n+1)(n+2)}{2}$$

Jest to całkowita liczba swobodnych parametrów wagowych, przykładowo:

- Dla n = 3, wektor $\dot{\mathbf{x}}$ ma 10 wymiarów
- Dla n = 10, wektor $\dot{\mathbf{x}}$ ma 65 wymiarów

- W przypadku funkcji dyskryminujących będących wielomianami m-tego rzędu, typowa $f_i(x)$ ma postać:

$$f_i(\mathbf{x}) = x_{i_1}^{e_1} x_{i_2}^{e_2} \dots x_{i_m}^{e_m}$$

$$\text{gdzie } 1 \le i_1, i_2, \dots, i_m \le n \text{ oraz } \mathrm{e}_\mathrm{i}, 1 \le i \le m \text{ to } 0 \text{ lub } 1.$$

  - Jest to wielomian, którego rząd jest z przedziału od 0 do m. Aby uniknąć powtórzeń wymagamy, aby $i_1 \le i_2 \le \dots \le i_m$

$$g^m(\mathbf{x}) = \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \dots \sum_{i_m=i_{m-1}}^{n} w_{i_1 i_2 \dots i_m} x_{i_1} x_{i_2} \dots x_{i_m} + g^{m-1}(\mathbf{x})$$

(gdzie $g^0(x) = w_{n+1}$) jest najogólniejszą funkcją dyskryminującą w postaci wielomianu rzędu m

Przykład 1: Niech *n = 3* i *m = 2,* to wówczas:

$$g^2(\mathbf{x}) = \sum_{i_1=1}^{3} \sum_{i_2=i_1}^{3} w_{i_1 i_2} x_{i_1} x_{i_2} + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4$$

$$= w_{11} x_1^2 + w_{12} x_1 x_2 + w_{13} x_1 x_3 + w_{22} x_2^2 + w_{23} x_2 x_3 + w_{33} x_3^2$$

$$+ w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4$$

Przykład 2: Niech *n = 2* i *m = 3,* to wówczas:

$$g^3(\mathbf{x}) = \sum_{i_1=1}^{2} \sum_{i_2=i_1}^{2} \sum_{i_3=i_2}^{2} w_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3} + g^2(\mathbf{x})$$

$$= w_{111} x_1^3 + w_{112} x_1^2 x_2 + w_{122} x_1 x_2^2 + w_{222} x_2^3 + g^2(x)$$

gdzie $g^2(\mathbf{x}) = \sum_{i_1=1}^{2} \sum_{i_2=i_1}^{2} w_{i_1 i_2} x_{i_1} x_{i_2} + g^1(\mathbf{x})$

$$= w_{11} x_1^2 + w_{12} x_1 x_2 + w_{22} x_2^2 + w_1 x_1 + w_2 x_2 + w_3$$

- Najczęściej używaną kwadratową funkcją dyskryminującą można reprezentować jako n-wymiarową powierzchnię kwadratową :

$$g(x) = x^T A x + x^T b + c$$

gdzie macierz $A = (a_{ij})$, wektor $b = (b_1, b_2, \ldots, b_n)^T$ i c, zależą od wag w_{ii}, w_{ij}, w_i z równania (2)

- Jeżeli A jest dodatnio określona to funkcja dyskryminująca jest hiperelipsoidą z osiami skierowanymi w kierunkach zgodnych z wektorami własnymi macierzy A
  - W szczególności: jeżeli $A = I_n$ (macierz jednostkowa), to funkcja dyskryminująca jest n-wymiarową hipersferą

- Jeżeli A jest ujemnie określona, to funkcja dyskryminująca opisuje hiperhiperboloidę

- Wniosek: macierz A w pełni determinuje kształt i charakterystykę funkcji dyskryminującej

Przykładowy problem: Rozważmy 3-wymiarową przestrzeń i funkcję dyskryminującą w postaci wielomianu trzeciego rzędu

1. Ile elementów zawiera wyrażenie reprezentujące funkcję dyskryminującą jeżeli tylko sześcienna i liniowa funkcja wchodzą w skład wyrażenia?

2. Przedstaw ogólne wyrażenie opisujące funkcję dyskryminującą w postaci wielomiany 4-tego rzędu dla 2-wymiarowej przestrzeni obrazów

3. Niech R^3 będzie oryginalną przestrzenią obrazów a funkcja dyskryminująca związana z klasami ω_1 i ω_2 ma postać:

$$g(\mathbf{x}) = 2x_1^2 + x_3^2 + x_2x_3 + 4x_1 - 2x_2 + 1$$

dla której $g(x) > 0$ *jeżeli* $x \in \omega_1$ i $g(x) < 0$ *jeżeli* $x \in \omega_2$

a) Zapisz $g(x)$ jako $g(x) = x^TAx + x^Tb + c$

b) Wyznacz klasy następujących obrazów:

(1,1,1), (1,10,0), (0,1/2,0)

- Macierze dodatnio określone

1. Kwadratowa macierz A jest *dodatnio określona* *jeżeli* $x^T Ax>0$ dla wszystkich niezerowych wektorów kolumnowych x.

2. Kwadratowa macierz A jest ujemnie określona jeżeli $x^T Ax < 0$ dla wszystkich niezerowych x.

3. Kwadratowa macierz A jest *dodatnio pół-określona* jeżeli $x^T Ax \geq 0$.

4. Kwadratowa macierz A jest *ujemnie pół-określona* jeżeli $x^T Ax \leq 0$ dla wszystkich x.

Powyższe warunki trudno jest sprawdzać bezpośrednio na podstawie definicji. W praktyce potrzebne są warunki mniej wymagające obliczeniowo.

Bardziej przydatne w praktyce są następujące własności, zachodzące wówczas gdy macierz A jest symetryczna.

i-ty minor główny macierzy A to macierz A_i utworzona z i pierwszych wierszy i kolumn macierzy A. Np., pierwszy główny minor macierzy A jest macierzą $A_i = (a_{11})$, drugi minor główny jest macierzą:

$$\mathbf{A}_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

i tak dalej…

- Macierz A jest dodatnio określona jeżeli wszystkie jej główne minory $A_1, A_2, \ldots, A_n$ mają wyznaczniki większe od zera

- Jeżeli te wyznaczniki są niezerowe a znaki kolejnych wyznaczników mają przeciwne znaki, począwszy od $det(A_1)<0$, to macierz A jest ujemnie określona

- Jeżeli wszystkie te wyznaczniki są nieujemne, to macierz jest dodatnio pół-określona

- Jeżeli znaki kolejnych wyznaczników mają przeciwne znaki, począwszy od $det(A_1)\leq 0$, to macierz jest ujemnie pół-określona

Dla przykładu, rozważmy macierz symetryczną o wymiarach 2x2:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \mathrm{a}_{12} \\ a_{21} & \mathrm{a}_{22} \end{pmatrix}$$

- Macierz jest dodatnio określona, jeżeli:
  a) $det(A_1) = a_{11} > 0$
  b) $det(A_2) = a_{11}a_{22} – a_{12}a_{12} > 0$

- Macierz jest ujemnie określona, jeżeli:
  a) $det(A_1) = a_{11} < 0$
  b) $det(A_2) = a_{11}a_{22} – a_{12}a_{12} > 0$

- Macierz jest dodatnio pół-określona, jeżeli:
  a) $det(A_1) = a_{11} \geq 0$
  b) $det(A_2) = a_{11}a_{22} – a_{12}a_{12} \geq 0$

- Macierz jest ujemnie pół-określona, jeżeli:
  a) $det(A_1) = a_{11} \leq 0$
  b) $det(A_2) = a_{11}a_{22} – a_{12}a_{12} \geq 0.$

Ćwiczenie 1: Sprawdź, czy poniższe macierze są dodatnio-, ujemnie- określone, dodatnio-, ujemnie- pół-określone lub nie należą do żadnej z tych grup.

$$(a)\ \mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$$

$$(b)\ \mathbf{A} = \begin{pmatrix} -2 & 4 \\ 4 & -8 \end{pmatrix}$$

$$(c)\ \mathbf{A} = \begin{pmatrix} -2 & 2 \\ 2 & -4 \end{pmatrix}$$

$$(d)\ \mathbf{A} = \begin{pmatrix} 2 & 4 \\ 4 & 3 \end{pmatrix}$$

Rozwiązania ćwiczenia 1:

- $A_1 = 2 > 0$
  $A_2 = 8 – 1 = 7 > 0$ $\Rightarrow$ A jest dodatnio określona
  positive definite

- $A_1 = -2$
  $A_2 = (-2 \times –8) –16 = 0$ $\Rightarrow$ A jest ujemnie pół-określona

- $A_1 = - 2$
  $A_2 = 8 – 4 = 4 > 0$ $\Rightarrow$ A jest ujemnie określona

- $A_1 = 2 > 0$
  $A_2 = 6 – 16 = -10 < 0$ $\Rightarrow$ A nie należy do żadnej z grup

Ćwiczenie 2:

Niech $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$

1. Wyznacz granice obszarów decyzyjnych odpowiadających macierzy A *($g(x) = x^T A x + x^T b + c$)* w przypadku, gdy $b^T = (1 , 2)$ i $c = - 3$

2. Rozwiąż $det(A-\lambda I) = 0$ i określ kształt i charakterystykę granicy obszarów decyzyjnych oddzielających dwie klasy ω_1 i ω_2

3. Przypisz klasy następującym punktom:
   - $x^T = (0 , - 1)$
   - $x^T = (1 , 1)$

Rozwiązanie ćwiczenia 2:

1.

$$g(\mathbf{x}) = (x_1, x_2)\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (x_1, x_2)\begin{pmatrix} 1 \\ 2 \end{pmatrix} - 3$$

$$= (2x_1 + x_2, x_1 + 4x_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + x_1 + 2x_2 - 3$$

$$= 2x_1^2 + x_1x_2 + x_1x_2 + 4x_2^2 + x_1 + 2x_2 - 3$$

$$= 2x_1^2 + 4x_2^2 + 2x_1x_2 + x_1 + 2x_2 - 3$$

2.

Dla $\lambda_1 = 3 + \sqrt{2}$ oraz $\begin{pmatrix} 2\text{-} & 1 \\ 1 & 4\text{-}\lambda \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$, otrzymujemy:

$$\begin{cases} (\text{-}1\text{-}\sqrt{2})x_1 + x_2 = 0 \\ x_1 + (1 - \sqrt{2})x_2 = 0 \end{cases} \Leftrightarrow (-1 - \sqrt{2})x_1 + x_2 = 0$$

Ostatnie równanie opisuje linię prostą współliniową z wektorem:

$$\vec{V}_1 = (1, 1 + \sqrt{2})^T$$

Dla $\lambda_2 = 3+\sqrt{2}$ oraz $\begin{pmatrix} 2\text{-}\lambda\lambda & 1 \\ 1 & 4\text{-}\lambda \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$, otrzymujemy:

$$\begin{cases} (\sqrt{2}-1)x_1 + x_2 = 0 \\ x_1 + (1+\sqrt{2})x_2 = 0 \end{cases} \Leftrightarrow (\sqrt{2}-1)x_1 + x_2 = 0$$

Ostatnie równanie opisuje linię prostą współliniową z wektorem:

$$\vec{V}_2 = (1, 1-\sqrt{2})^T$$

Elipsa będąca granicą obszarów decyzyjnych ma dwie osie współliniowe odpowiednio z wektorami V_1 i V_2

3. $X = (0\,,\,-1)^T \Rightarrow g(0\,,\,-1) = -1 < 0 \Rightarrow x \in \omega_2$

$X = (1\,,\,1)^T \Rightarrow g(1\,,\,1) = 8 > 0 \Rightarrow x \in \omega_1$