

www.kwmimkm.polisi.pl

# METODY HEURYSTYCZNE

## wykład 5

www.kwmimkm.polisi.pl

# SSN

## cd.

www.kwmimkm.polisi.pl

### KLASY ZASTOSOWAŃ:

**PREDYKCJA**

ze znajomości:  
przewidzieć :

bez jawnego definiowania związku między danymi wejściowymi a wyjściowymi

www.kwmimkm.polisi.pl

### KLASYFIKACJA I ROZPOZNAWANIE WZORCÓW

Zaszeregowanie danych wejściowych do jednej z klas:

np. sieć pozwala na podstawie danych bilansowych stwierdzić, czy dane przedsiębiorstwo należy do zwyżkujących gospodarczo, czy przeżywa stagnację czy też grozi mu regres.

www.kwmimkm.polisi.pl

### APROKSYMACJA (interpolacja, ekstrapolacja)

ze znajomości:  
odtworzyć:

**ASOCJACJA**

Podanie danego wzorca na wejście powinno powodować pojawienie się odpowiadającego mu wzorca na wyjściu.

www.kwmimkm.polisi.pl

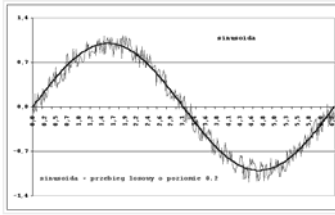
### STEROWANIE

Dobrać  $f(k)$  tak aby  $\|x(k) - u(k)\| \rightarrow 0, k \rightarrow \infty$

### KOJARZENIE DANYCH

automatyzacja procesów wnioskowania i wykrywanie istotnych powiązań między danymi.

## FILTRACJA SYGNAŁÓW



## OPTYMALIZACJA

statyczna i dynamiczna, optymalizacja kombinatoryczna i zagadnienia bardzo trudne obliczeniowo.



7

## Dobór optymalnej struktury SSN:

3 główne grupy algorytmów pomocnych w utworzeniu sieci optymalnej :

- metody wzrostu;
- metody redukcji,
- metody optymalizacji dyskretnej.

## Metody wzrostu:

- na początku procesu optymalizacji struktura sieci powinna być możliwie mała;
- w kolejnych iteracjach są dodawane kolejne neurony ukryte (co powinno powodować zwiększenie sprawności działania sieci);
- neurony są dodawane aż do osiągnięcia punktu optymalnego.



8

## Metody redukcji

- na początku procesu optymalizacji struktura sieci powinna być możliwie złożona;
- w kolejnych iteracjach są usuwane kolejne neurony lub połączenia między nimi (co powinno powodować zwiększenie sprawności działania sieci);
- postępowanie jest powtarzane aż do osiągnięcia punktu optymalnego.



9

## Metody optymalizacji dyskretnej

- opierają się na założeniu, że proces nauki sieci i wyboru architektury zachodzą równocześnie;
- czynnikiem ocenianym jest określona funkcja, reprezentująca jakość danej sieci;
- w kolejnych krokach sieci dobierane są tak, by dążyć do maksymalizacji funkcji jakości;
- możliwe jest wykorzystanie AE jako metody optymalizacji.

Żadna z tych metod **nie jest idealna**. Często wybór którejś z nich zależy od **rodzaju rozwiązywanego problemu**.



10

- **Zwykle:** uczenie z początkowo małą liczbą neuronów i stopniowe zwiększanie ich liczby.

- Zbyt mało neuronów w warstwie ukrytej – sieć nie potrafi poprawnie odwzorować funkcji.

- Zbyt wiele elementów warstwy ukrytej:

- wydłużenie procesu uczenia;
- uczenie się „na pamięć” (szczególnie, gdy liczba próbek w ciągu uczącym jest niewielka) - sieć poprawnie rozpoznaje tylko sygnały zgodne z tymi w ciągu uczącym (brak generalizacji przy dobrej interpolacji).



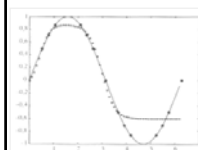
11

Np. (L. Rutkowski, *Metody i techniki sztucznej inteligencji*, PWN, W-wa 2006):

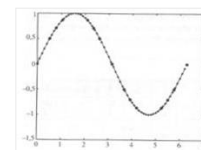
$$y = \sin(x), \quad x \in [0, 2\pi]$$

Ciąg uczący:

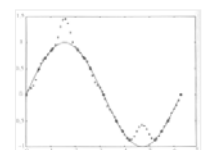
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Wejście $x$	0	$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\pi}{4}$	$\pi$	$2\pi$	$\frac{7\pi}{6}$	$\frac{4\pi}{3}$	$\frac{5\pi}{4}$	$\frac{5\pi}{6}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{3}$	$\frac{11\pi}{6}$	$\frac{7\pi}{4}$
Oczekiwane wyjście $d=f(x)$	0	0.5	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	0	0	-0.5	$-\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{3}}{2}$	$-\frac{1}{2}$	$-\frac{\sqrt{2}}{2}$



1-2-1



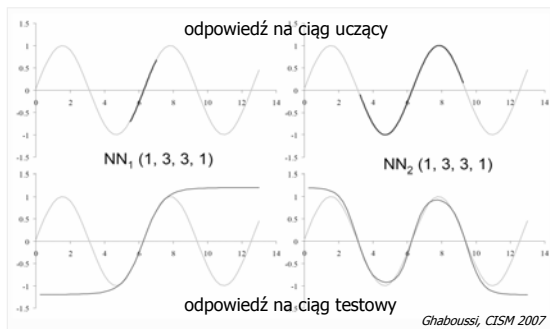
1-3-1



1-15-1

12

Dobór próbek (ciąg uczący) wpływa na jakość nauczania:



13

## JAK DŁUGO UCZYĆ SIEĆ?

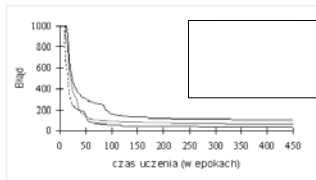
**Liczba prezentacji** ciągu uczącego konieczna do nauczenia sieci:

- prognozowanie finansowe:  $10^9$
- synteza mowy:  $10^{10}$
- rozpoznawanie mowy lub pisma odręcznego:  $10^{12}$
- rozpoznawanie znaków Kenji:  $10^{13}$

14

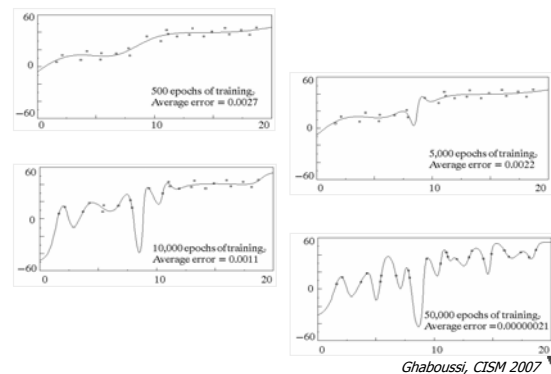
- Uważa się, iż **czas uczenia** sieci rośnie **wykładniczo** wraz ze wzrostem liczby elementów sieci.
- Korzystne jest pokazywanie elementów ciągu uczącego **w różnej kolejności**.

**Malenie błęd** jest różne dla różnie wylosowanych wag początkowych (dla tej samej sieci):



15

Zbyt długie uczenie również może skutkować utratą zdolności uogólniania:



16

## UCZENIE WARSTW UKRYTYCH

Bezpośrednie wyznaczenie błędów nie jest możliwe (sygnałów WY z warstwy ukrytej nie ma z czym porównać).

**Metoda wstecznej propagacji błędów** (backpropagation):

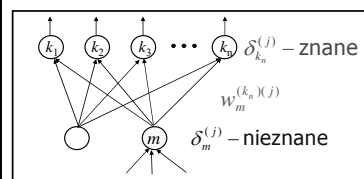
- Zmiana wagi – jak przy sieci jednowarstwowej;
- Obliczanie  $\delta$  – sumowanie błędów z następnej warstwy

Błędy w warstwach ukrytych są wyznaczone **w sposób przybliżony**.

17

$$\delta_m^{(j)} = \sum_{k=1}^n w_m^{(k)(j)} \delta_k^{(j)}$$

- $m$  – numer neuronu w warstwie ukrytej
- $n$  – liczba neuronów w warstwie następnej  $k$ ;
- $j$  – numer kroku uczenia;
- $\delta_m^{(j)}$  – błąd popełniany przez neuron  $m$ ;
- $\delta_k^{(j)}$  – błąd popełniany przez neuron w warstwie  $k$



- obliczanie błędów w warstwie WY;
- obliczanie błędów w warstwie poprzedniej;
- itd. aż do warstwy pierwszej.

18

## UCZENIE SIECI NEURONOWYCH

- Zamiast programowania!
- Wymuszanie określonego reagowania sieci na zadane sygnały wejściowe (*poprzez odpowiedni dobór wag*).

**Ta sama sieć** może służyć do rozwiązywania skrajnie różnych zadań.

### Warianty uczenia:

- uczenie bez nauczyciela (nienadzorowane);
- uczenie z nauczycielem (nadzorowane);
- uczenie z krytykiem.



19

## Uczenie bez nauczyciela (*unsupervised learning*)

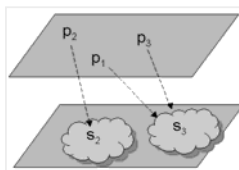
- **Donald Hebb** (fizjolog i psycholog) – w umyśle zachodzą procesy wzmacniania połączeń między neuronami, jeśli zostały one pobudzone jednocześnie.
- Pożądana odpowiedź nie jest znana.
- Sieć uczy się poprzez analizę reakcji na pobudzenia; samoorganizacja struktury – wszelkie *regularności*, *linie podziału* i inne *charakterystyki* danych wejściowych sieć musi wykryć sama.
- Zdolności do wykrywania skupisk obrazów wejściowych są wykorzystywane do ich klasyfikacji, gdy klasy nie są z góry ustalone.



20

## Uczenie bez nauczyciela (*unsupervised learning*)

- Sieci pokazuje się kolejne przykłady bez określenia, co trzeba z nimi zrobić.
- Różne pobudzenie różnych neuronów - połączenia między źródłami silnych sygnałów a neuronami, które na nie reagują są wzmacniane.
- W sieci stopniowo powstają wzorce poszczególnych typów sygnałów rozpoznawane przez pewną część neuronów.
- Uczenie spontaniczne, odkrywanie ciekawych struktur w przestrzeni danych, korelacja zachowań systemu ze zmianą tych struktur – dominuje w okresie niemowlęcym.



21

## Uczenie bez nauczyciela - wady

- Zwykle powolniejsze.
- Nie wiadomo, który neuron będzie rozpoznawał jaki sygnał.
- Część sygnałów może być rozpoznawana przez więcej niż jeden neuron.
- Część sygnałów może nie być rozpoznawana przez żaden neuron (sieć musi być większa niż przy nauczycielu zwykle przynajmniej 3 razy).



22

## Uczenie z nauczycielem (*supervised learning*)

- Podawanie sieci zestawów sygnałów WE wraz z prawidłowym sygnałem WY.
- Naśladowanie nauczyciela, jakim jest ciąg uczący (podejście „szkolne”).
- Zestawy sygnałów (zwykle) powtarza się wielokrotnie, zaś sieć modyfikuje wagi na wejściach tak, by zminimalizować błąd.
- Zmiana wagi na  $i$ -tym wejściu neuronu po pokazaniu  $j$ -ego obiektu uczącego jest proporcjonalna do popełnianego na tym etapie błędu  $\delta^{(j)}$ .



23

## Uczenie z krytykiem (*reinforcement learning*):

- Odmiana uczenia nadzorowanego.
- Nauczyciel nie dysponuje pełną wiedzą na temat wszystkich prawidłowych odpowiedzi.
- Zamiast informacji o pożądanym WY, sieć dysponuje jedynie oceną efektu swego działania w ramach dwóch prostych kategorii.
- Ocena  $\Rightarrow$  wzmocnienie (pozytywne lub negatywne)  $\Rightarrow$  odpowiednie zmiany wag.



24

## Uczenie z krytykiem (*reinforcement learning*):

- Optymalizacja zysków na dłuższą metę.
- Np.: gry z przeciwnikiem, krytyką jest przegrana lub wygrana na końcu partii.
- Uczenie z krytykiem lub z „wzmocnieniem” pożądanych zachowań po dłuższym okresie.
- Uczenie dojrzałe (nabieranie „mądrości”).
- Bardziej uniwersalne w zastosowaniu podejście do problemu.
- Praktyczna realizacja jest bardziej skomplikowana.



25

## SAMOUCZENIE SIECI

- Cała wiedza, jaką sieć może zdobyć jest zawarta w obiektach pokazywanych (muszą zawierać klasy podobieństwa).
- Nie mogą to być obiekty całkiem przypadkowe, ale tworzyć skupiska wokół pewnych ośrodków.
- Proces samouczenia utrwała i pogłębia wrodzone zdolności neuronów.



26

## SAMOUCZENIE SIECI:

Ciąg uczący:  $U = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(N)}\}$

- $\mathbf{X}^{(j)}$  –  $n$ -wymiarowy wektor danych wejściowych w  $j$ -ym kroku uczenia
- $N$  – liczba posiadanych pokazów.

Reguła uczenia dla  $m$ -tego neuronu w  $j$ -ym kroku:

$$w_i^{(m)(j+1)} = w_i^{(m)(j)} + \eta x_i^{(m)} y_m^{(j)}$$

gdzie:

$$y_m^{(j)} = \sum_{i=1}^n w_i^{(m)(j)} x_i^{(j)}$$



27

$$w_i^{(m)(j+1)} = w_i^{(m)(j)} + \eta x_i^{(m)} y_m^{(j)}$$

$$y_m^{(j)} = \sum_{i=1}^n w_i^{(m)(j)} x_i^{(j)}$$

- Wielkość zmiany wagi – liczona na podstawie iloczynu sygnału na odpowiednim wejściu przez sygnał wyjściowy neuronu.
- Uczenie to zwie się też korelacyjnym – zmierza do takiego dopasowania wag, by uzyskać najlepszą korelację między sygnałami WE a zapamiętanym (w formie wag) wzorcem sygnału, na który dany neuron ma reagować.



28

## KONKURENCJA W SIECIACH SAMOUCZĄCYCH

WTA (Winner Takes All) - zwycięzca bierze wszystko:

- Najlepszy neuron – **niezerowa** wartość sygnału WY (zwykle 1);
- Pozostałe wyjścia są **zerowane**;
- Tylko **najlepszy** neuron jest uczony.

☞ Zasada WTA daje jednoznaczną odpowiedź sieci (co niekoniecznie musi być zaletą).

☞ Wszystkie sygnały mniejsze niż ustalony próg – brak rozpoznania.



29

## Efekty:

- Neuron, który raz wygrał przy pokazaniu danego wzorca – dalej będzie wygrywał.
- Samouczenie jest skuteczniejsze i efektywniejsze (*każdy neuron rozpoznaje jeden obiekt, pozostałe neurony pozostają niezagospodarowane*).
- Wysoce prawdopodobne jest, iż:
  - nie będzie grup neuronów rozpoznających ten sam wzorec;
  - nie będzie klas nierozpoznanych przez żaden neuron.
- Niezagospodarowane neurony są gotowe rozpoznawać nowe wzorce.
- Po wykorzystaniu wszystkich neuronów i pojawieniu się nowego wzorca – przeciąganie któregoś z neuronów w jego stronę.



30

## SIECI SAMOORGANIZUJĄCE SIĘ (Teuvo Kohonen)

- Tworzą odwzorowania sygnałów WE w sygnały WY, spełniające pewne ogólne kryteria (nie zdeterminowane przez twórcę ani użytkownika sieci) – samoorganizacja sieci.
- Inny (wyższy) sposób samouczenia, wprowadzający efekty koherencji i kolektywności.

**Koherencja** (słown. spójność, spoiłość, łączność) - sieć grupuje dane wejściowe wg wzajemnego podobieństwa – wykrywa automatycznie obiekty podobne do siebie nawzajem i inne od innych grup obiektów.



31

**Analiza skupień** – w analizie ekonomicznej (np. podobieństwo przedsiębiorstw – rentowność).

**Kolektywność** – to co rozpoznaje neuron zależy w dużej mierze od tego, co rozpoznają inne neurony.

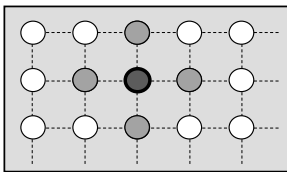
**Sąsiedztwo** – znaczenie ma wzajemne położenie neuronów w warstwach.

Zwykle sąsiedztwo 2-wymiarowe – neurony w węzłach regularnej siatki (każdy neuron ma min. 4 sąsiadów).

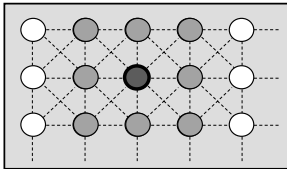


32

### Przykładowe sąsiedztwa:

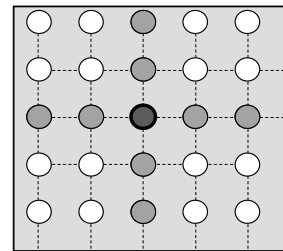


Rozbudowane sąsiedztwo:



33

### Rozszerzone sąsiedztwo:

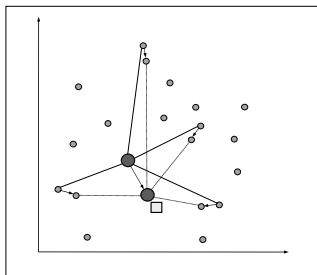


Jednowymiarowe sąsiedztwo:



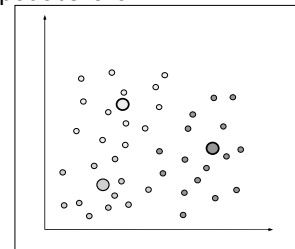
34

Gdy w procesie uczenia któryś neuron zwycięża: uczy się też (w mniejszym stopniu) jego sąsiadów (niezależnie od ich wag początkowych!).



35

Neurony sąsiadujące rozpoznają sygnały z sąsiadujących podobszarów:



**Skutek:**

Sygnały równomiernie rozmieszczone w pewnym obszarze  $\Rightarrow$  neurony zostają tak nauczone, by każdy podobszar sygnałów był rozpoznawany przez inny neuron.



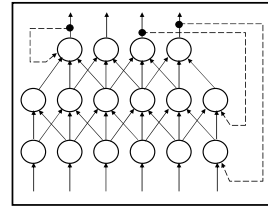
36

- Po wytrenowaniu każdej praktycznej sytuacji odpowiada neuron, który ją reprezentuje.
  - Sąsiedztwo powoduje **wykrywanie sytuacji podobnych** do prezentowanych.
  - W sieci powstaje **wewnętrzny obraz świata zewnętrznego**.
  - Sygnały blisko siebie będą wykrywane przez **leżące blisko siebie** neurony.
- Przykładowe zastosowania:**
- **robot** dostosowujący zachowanie do zmiennego środowiska;
  - **systemy bankowe** – stworzenie modelu wiarygodnego kredytobiorcy.



## SIECI REKURENCYJNE

Zawierają sprzężenia zwrotne:



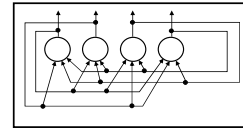
Po jednorazowym podaniu sygnału WE – **długotrwały proces** zmiany sygnału WY, w efekcie **stan równowagi**.



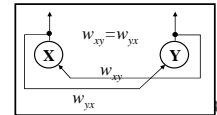
- Waga sprzężenia dodatnia – sygnał zmienia się **jednokierunkowo** (aperiodycznie);
  - Waga sprzężenia ujemna – sygnał zmienia się **oscylacyjnie**;
  - Przy neuronach nieliniowych możliwe chaotyczne błędzenie sygnałów;
- Zachowania:**
- stabilne (*zbieganie się sygnałów do określonej wartości*);
  - niestabilne (*wartości sygnału coraz większe*).
- Zastosowania:**
- zadania optymalizacji (*stany równowagi odpowiadają rozwiązaniom zadań*);
  - pamięci skojarzeniowe (*drobny fragment informacji pozwala odtworzyć całą informację*).



## SIECI HOPFIELDA



- Każdy neuron jest związany z każdym innym na zasadzie obustronnego sprzężenia zwrotnego.
- Zabroniono sprzężeń zwrotnych obejmujących pojedynczy neuron.
- Symetria współczynników wagowych:



**Skutek:** zachodzące procesy są **zawsze stabilne**.



# ZBIORY ROZMYTE I WNIOSKOWANIE PRZYBLIŻONE



1965 – Lotfi A. Zadeh: „Fuzzy sets”



“In almost every case you can build the same product without fuzzy logic, but fuzzy is faster and cheaper.”  
Prof. Lotfi Zadeh, UC Berkeley, Inventor of Fuzzy Logic



Metoda reprezentacji wiedzy wyrażonej w języku naturalnym:

- „Temperatura wynosi 29°C” – informacja liczbową - naturalna dla systemów komputerowych.
- „Jest dość ciepło” – informacja opisowa - naturalna dla człowieka.

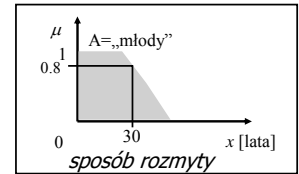
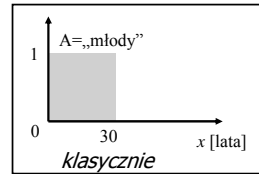
**Klasyczna teoria zbiorów:** dowolny element należy lub nie należy do danego zbioru.

**Teoria zbiorów rozmytych:** element może częściowo należeć do pewnego zbioru.



Zamiast dwóch wartości logicznych (prawda i fałsz) nieskończenie wiele wartości [0,1].

Np.: „młody człowiek”:



Umożliwiają formalne określenie pojęć nieprecyzyjnych i wieloznacznych:

- „wysoki hałas”,
- „małe zarobki”,
- „niskie zużycie paliwa”.



**Obszar rozważań X** (*the universe of discourse*) - zbiór nierozmyty (np. płaca w UK i w Polsce).

**Zbiór rozmyty** w pewnej przestrzeni (niepustej) X - zbiór par:

$$A = \{(x, \mu_A(x)); x \in X\}$$

$\mu_A(x)$  – funkcja przynależności zbioru rozmytego A.

**Funkcja przynależności** – przypisuje każdemu elementowi  $x \in X$  stopień jego przynależności do zbioru rozmytego A



- $\mu_A(x) = 1$  – pełna przynależność elementu  $x$  do ZR A;
- $\mu_A(x) = 0$  – brak przynależności  $x$  do ZR A;
- $0 < \mu_A(x) < 1$  – częściowa przynależność  $x$  do ZR A.

Stopień przynależności to **nie jest** prawdopodobieństwo: młody w 80% to nie 4 młodych na 5...

Symboliczny zapis ZR o skończonej liczbie elementów:

$$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \dots + \frac{\mu_A(x_n)}{x_n} = \sum_{i=1}^n \frac{\mu_A(x_i)}{x_i}$$

suma mnogościowa                      przyporządkowanie



**Np.** „Ciepła woda na basenie”:

- Obszar rozważań:  $X = [15, 21, \dots, 35]$
- Zbiór rozmyty A (według osoby nr 1):

$$A = \frac{0.1}{20} + \frac{0.3}{21} + \frac{0.4}{22} + \frac{0.6}{23} + \frac{0.8}{24} + \frac{1}{25} + \frac{0.9}{26} + \frac{0.8}{27} + \frac{0.75}{28} + \frac{0.7}{29}$$

- Według osoby nr 2:

$$A = \frac{0.2}{18} + \frac{0.4}{19} + \frac{0.6}{20} + \frac{0.8}{21} + \frac{1}{22} + \frac{0.8}{23} + \frac{0.6}{24} + \frac{0.4}{25} + \frac{0.2}{26}$$

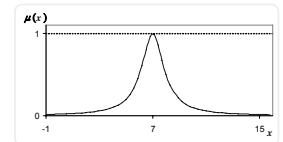
Jeśli X - przestrzeń o nieskończonej liczbie elementów, to zapis symboliczny:

$$A = \int_x \frac{\mu_A(x)}{x}$$

**Np.** „Zbiór liczb bliskich liczbie 7”:

$$\mu_A(x) = \frac{1}{1 + (x - 7)^2}$$

$$A = \int_x \frac{[1 + (x - 7)^2]^{-1}}{x}$$



lub

$$\mu_A(x) = \begin{cases} 1 - \frac{\sqrt{|x-7|}}{3} & \text{jeżeli } 4 \leq x \leq 10 \\ 0 & \text{w przeciwnym razie} \end{cases}$$

