

SIECI RBF (RADIAL BASIS FUNCTIONS)

Wybrane slajdy z prezentacji prof. Tadeusiewicza

Wykład Andrzeja Burdy

S. Osowski, “Sieci Neuronowe w ujęciu algorytmicznym”,
Rozdz. 5, PWNT, Warszawa 1996.

opr. P.Lula, R.Tadeusiewicz, STATISTICA Neural Networks
PL. Przewodnik problemowy, StatSoft, Kraków 2001

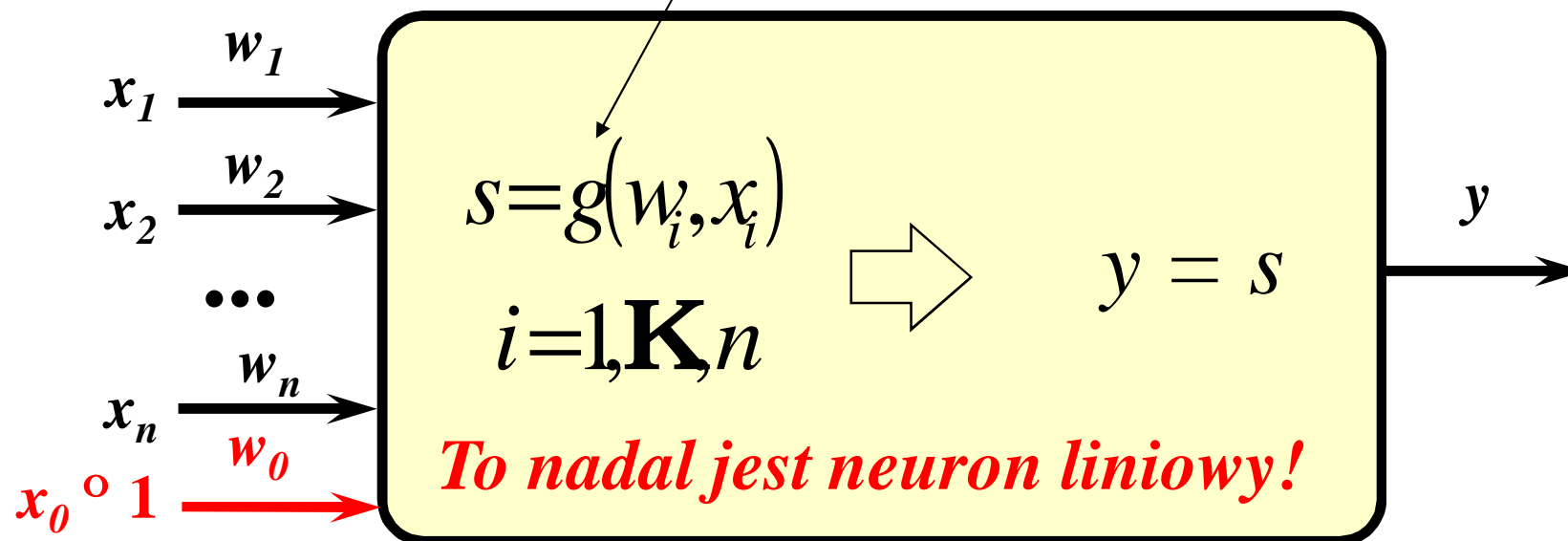
http://www.statsoft.pl/textbook/stathome_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstneunet.html

Joanna Grabska-Chrząstowska

Z neuronem liniowym (i z innymi neuronami budowanymi na jego bazie) związana jest jeszcze sprawa wyrazu wolnego w formule agregacji

Czysta agregacja liniowa: $S = \sum_{i=1}^n w_i x_i$

ma wadę, polegającą na tym, że charakterystyka neuronu **musi** tu przechodzić przez początek układu



Żeby zachować liniową postać wzoru opisującego neuron dodaje się dodatkowe **pseudo-wejście** nazywane BIAS, które zawsze dostarcza sygnał 1

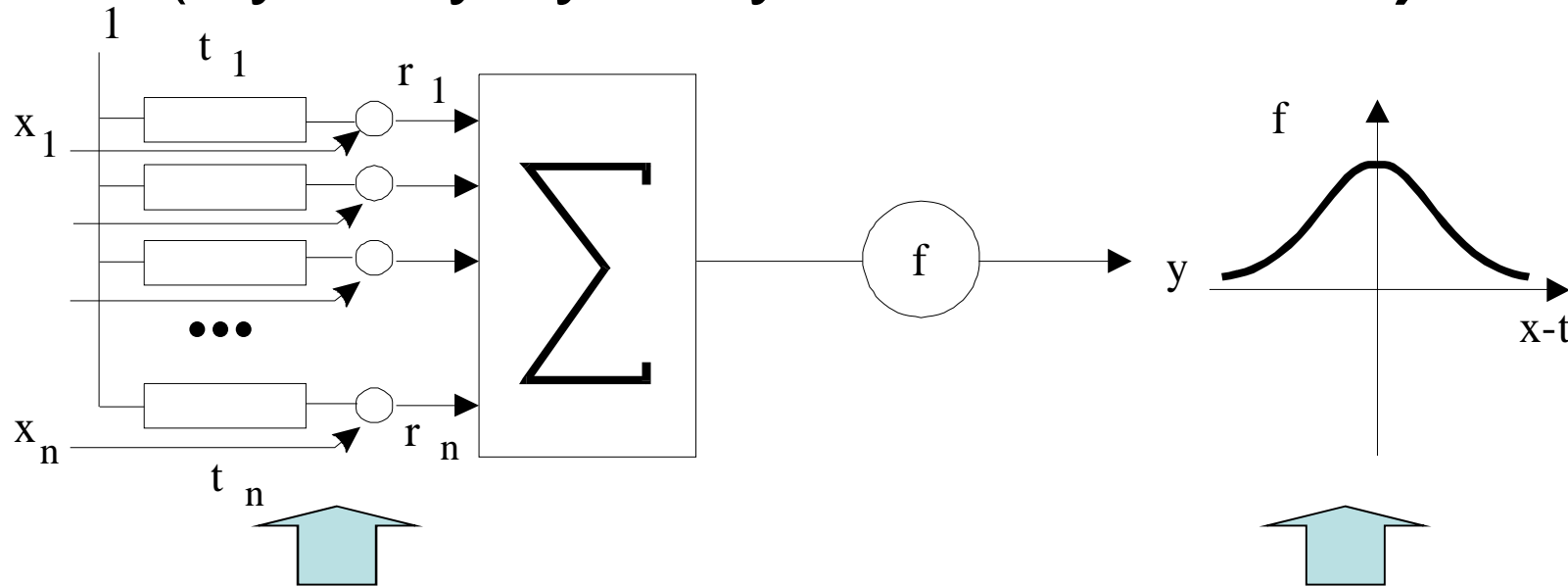
Bogatsze możliwości daje **agregacja afiniczna** (z wyrazem wolnym w formule):

$$s = \sum_{i=1}^n w_i x_i + w_0$$

Wtedy agregacja **jest nadal** liniowa:

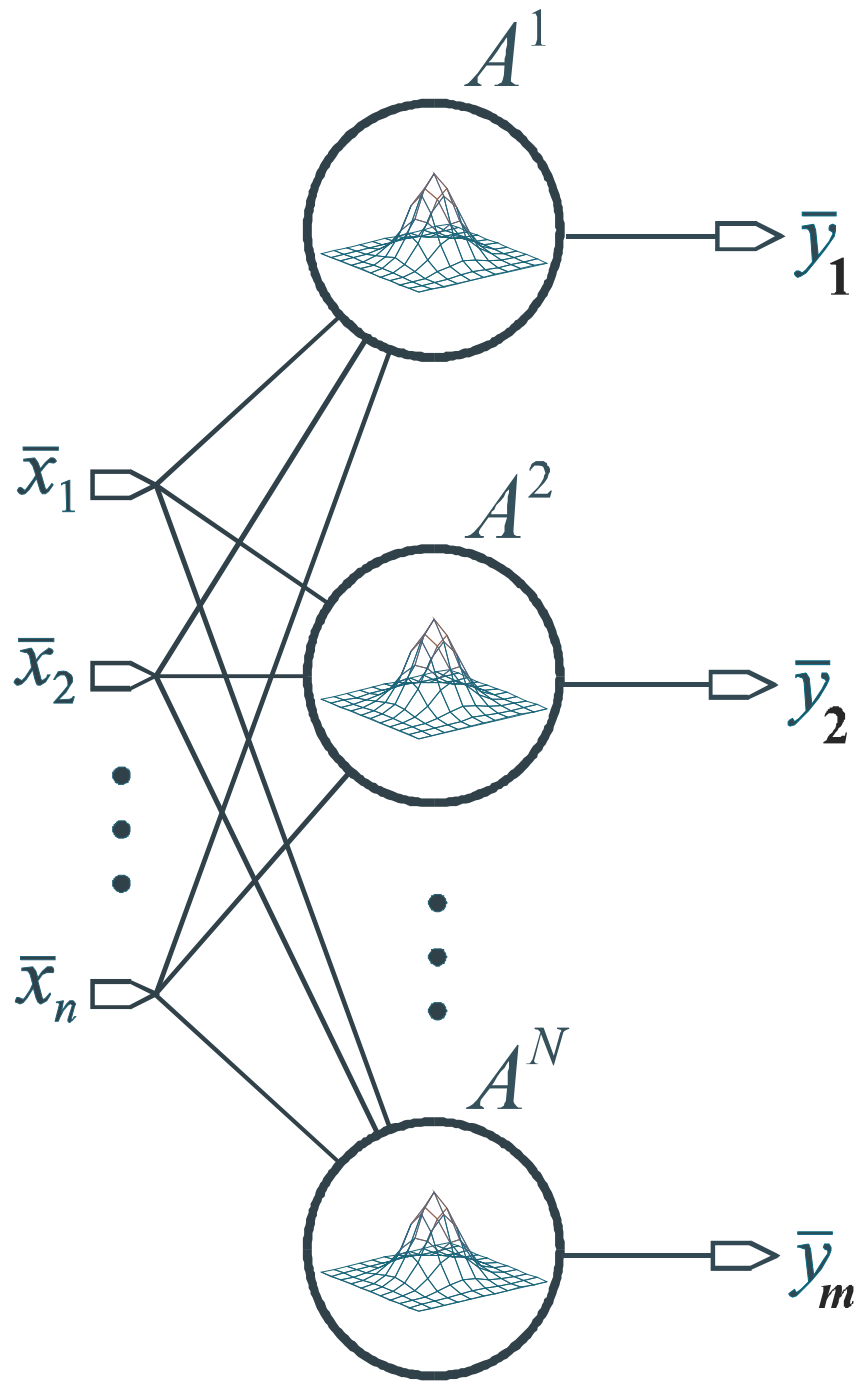
$$s = \sum_{i=0}^n w_i x_i$$

Odmiennie działającym elementem używanym w niektórych typach jest tzw. neuron radialny (wykorzystywany w sieciach **RBF**)



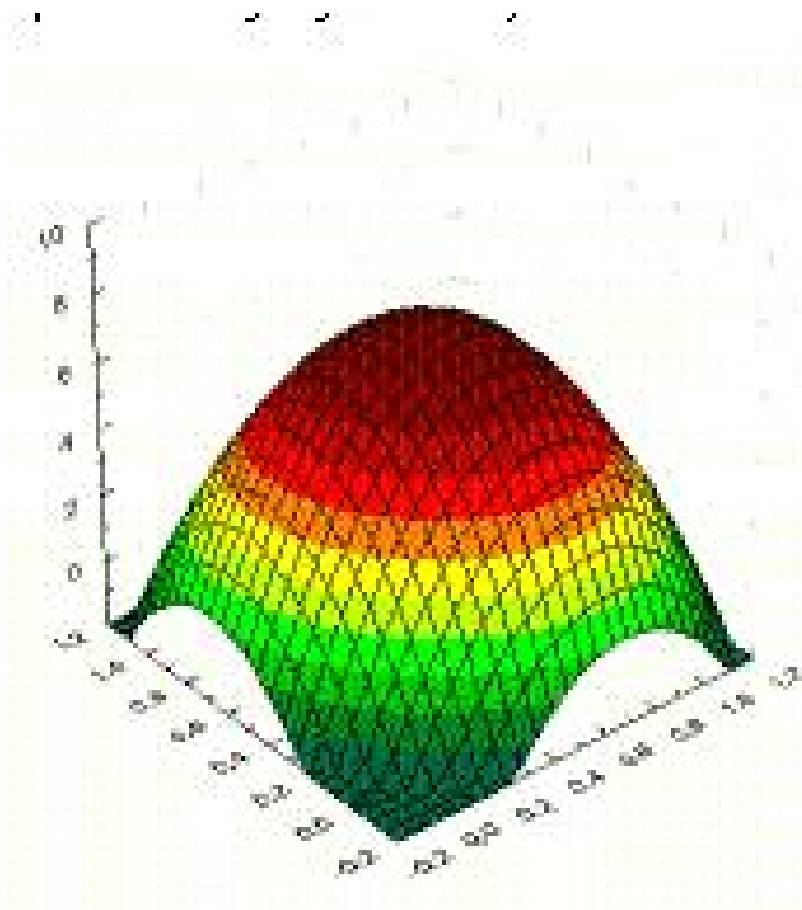
Agregacja sygnałów wejściowych w tym typie neuronu polega na obliczaniu **odległości** pomiędzy obecnym wektorem wejściowym X a ustalonym podczas uczenia **centroidem** pewnego podzbioru T

Również nieliniowa funkcja przejścia w tych neuronach ma odmienną formę - „dzwonu” gaussoidy - czyli jest funkcją **niemonotoniczną**.



Sieć typu RBF
w zastosowaniu
do klasyfikacji
(wykrywa
i sygnalizuje
skupiska
danych
wejściowych)

NEURON RADIALNY



Neuron radialny jest zdefiniowany przez swoje **centrum** oraz parametr określany jako "**promień**". Punkt w przestrzeni N -wymiarowej jest definiowany przy użyciu N liczb, co dokładnie odpowiada liczbie wag w neuronie liniowym, z tego powodu centrum neuronu radialnego jest przechowywane w zestawie parametrów określanych w programie *STATISTICA* również jako "**wagi**" (choć przy wyznaczaniu łącznego pobudzenia neuronów nie dokonuje się mnożenia składowych sygnałów przez te "wagi" tylko wyznaczana jest odległość wektora wag i wektora sygnałów wejściowych). Promień (lub inaczej odchylenie) jest przechowywany w neuronie jako tak zwana "**wartość progowa**".

WYBÓR PARAMETRÓW CENTRÓW

Powtórne próbkowanie. Metoda ta polega na tym, że wybrane w sposób losowy elementy ze zbioru uczącego kopiowane są do neuronów radialnych (jako występujące w tych neuronach zestawy wag). Ponieważ podlegające kopiowaniu sygnały wejściowe wybrane zostały w sposób losowy, więc "reprezentują" one (w sensie statystycznym) rozkład wszystkich danych uczących. Jednakże, jeśli liczba neuronów radialnych nie jest duża, to neurony te mogą stanowić w rzeczywistości złą reprezentację (Haykin, 1994).

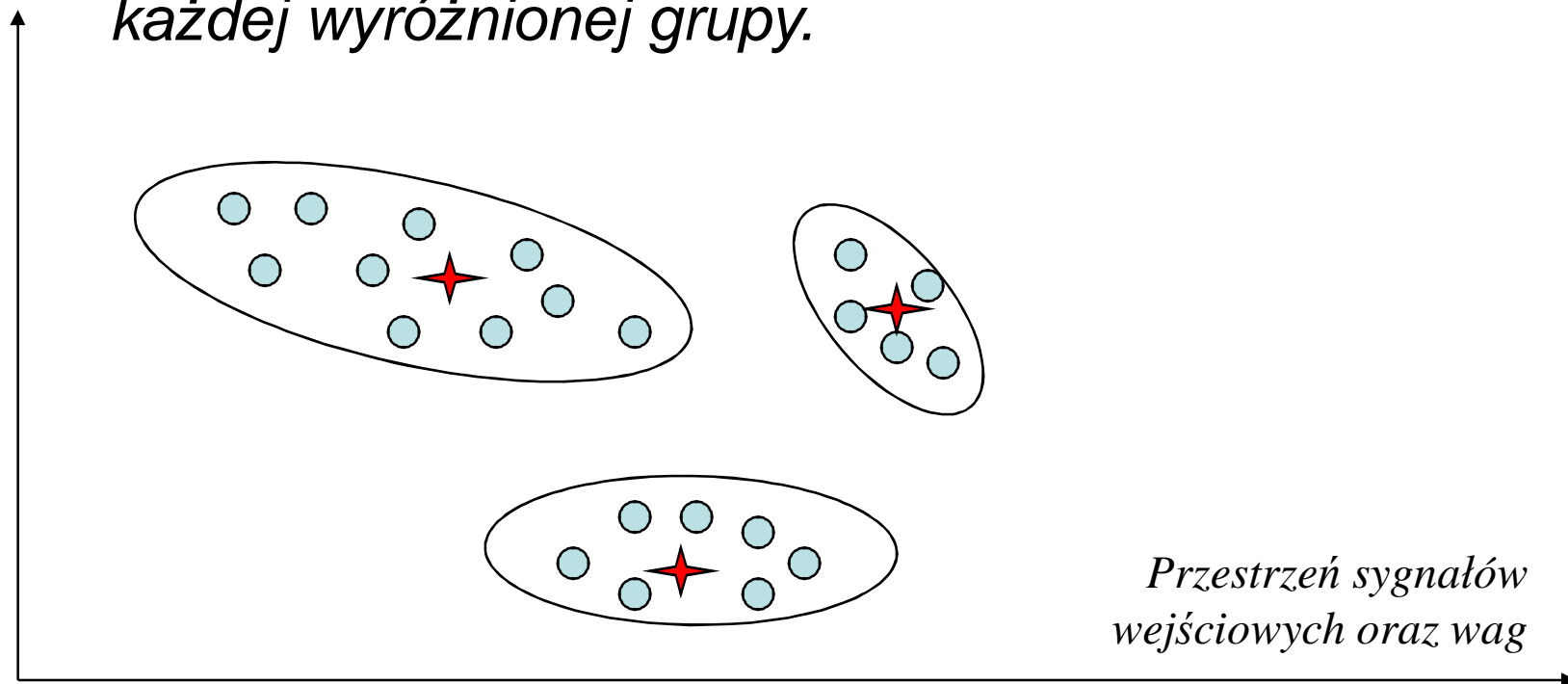
Algorytm k-średnich. Algorytm ten (Bishop, 1995) próbuje ustalić optymalny zbiór punktów, które stanowią centra skupień występujących w danych uczących. Mając K neuronów radialnych musimy wytworzyć K niezależnych centrów, w taki sposób, by reprezentowały one charakterystyczne skupiska wejściowych danych. Centra te ustala się w procesie iteracyjnym, w którym powtarzane są następujące czynności

- Ø Każdy element zbioru uczącego przypisywany jest do tego centrum skupienia, które położone jest bliżej danego elementu niż wszystkie pozostałe centra;
- Ø Każde centrum skupienia wyznaczone jest jako wektor średnich wartości zmiennych wyznaczony dla wszystkich punktów należących do danego skupienia.

Określenie wag neuronów radialnych metodą K-średnich

Elementy zbioru uczącego dzielone są na grupy elementów podobnych (metodą k-średnich).

W charakterze wag stosowane są środki ciężkości każdej wyróżnionej grupy.



WYBÓR PROMIENIA (ODCHYLENIA)

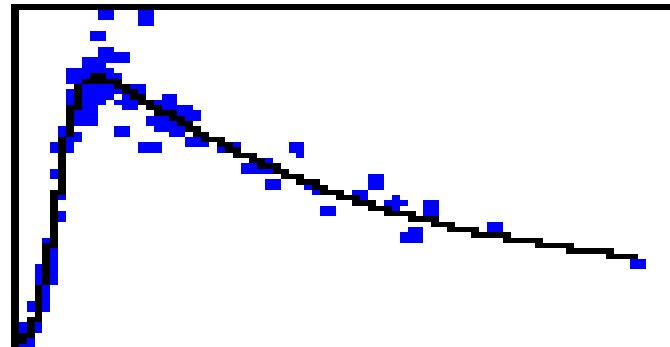
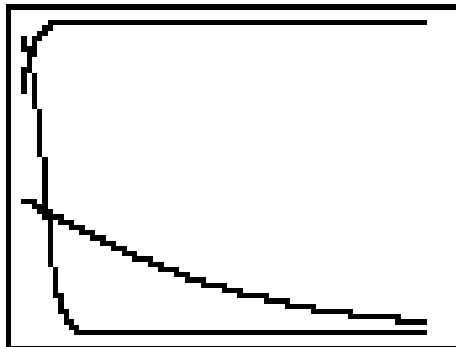
Definiowanie przez użytkownika. Użytkownik samodzielnie określa wielkość odchylenia.

Równomierny przydział odchyleń. Odchylenie (identyczne dla wszystkich neuronów) jest określane za pomocą pewnej reguły heurystycznej, uwzględniającej liczbę centrów oraz wielkość zajmowanej przez nie przestrzeni (Haykin, 1994).

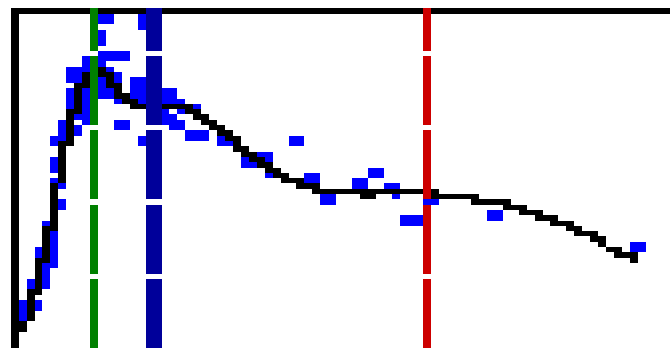
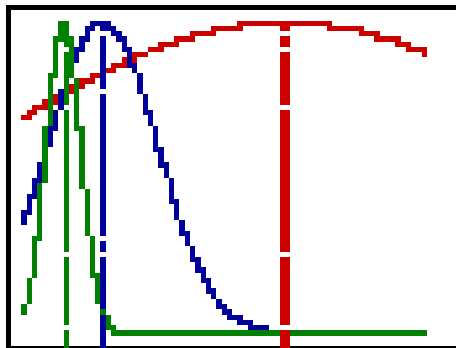
Przydział metodą k-najbliższych sąsiadów. Odchylenie dla każdego neuronu jest określane indywidualnie jako średnia odległość do jego k najbliższych sąsiadów (przypadków ze zbioru danych). Stąd odchylenia są mniejsze w mocno zagęszczonym obszarze danych, co umożliwia zachowanie drobnych szczegółów i większe w obszarze, w którym dane występują rzadko (umożliwia to lepszą interpolację).

Zastosowanie **RBF** (zamiast **MLP**) spowoduje, że sieć neuronowa znajdzie aproksymację lepiej dopasowaną do lokalnych właściwości zbioru danych, ale gorzej ekstrapolującą.

MLP



RBF



Funkcja bazowe

Wynik dopasowania

Network Illustrations: Iris5FN.vla

Index	Profile	Train Part.	Select Part.	Test Part.	Train Error
1	MLP 4-4-4-3-1	0.973684	1.000000	0.998919	0.201139
2	FBF 4-4-8-3-1	0.980526	0.945946	0.887822	0.248233
3	PHN 4-4-76-3-1	0.947368	0.972973	0.891992	0.176084
4	SDFM 4-4-12-1	0.000000	0.000000	0.000000	0.159535
5	GFNN 4-4-76-4-...	0.990526	0.854855	0.787794	0.180450
6	Linear 4-4-3-1	0.004737	0.700704	0.700704	0.275002
7	PCA 4-4-4-4	0.000000	0.000000	0.000000	0.000000
8	Class 4-4-4-4	0.000000	0.700704	0.700704	0.149904

Illustration Custom Case

Cancel

Options

Select models

Network graph

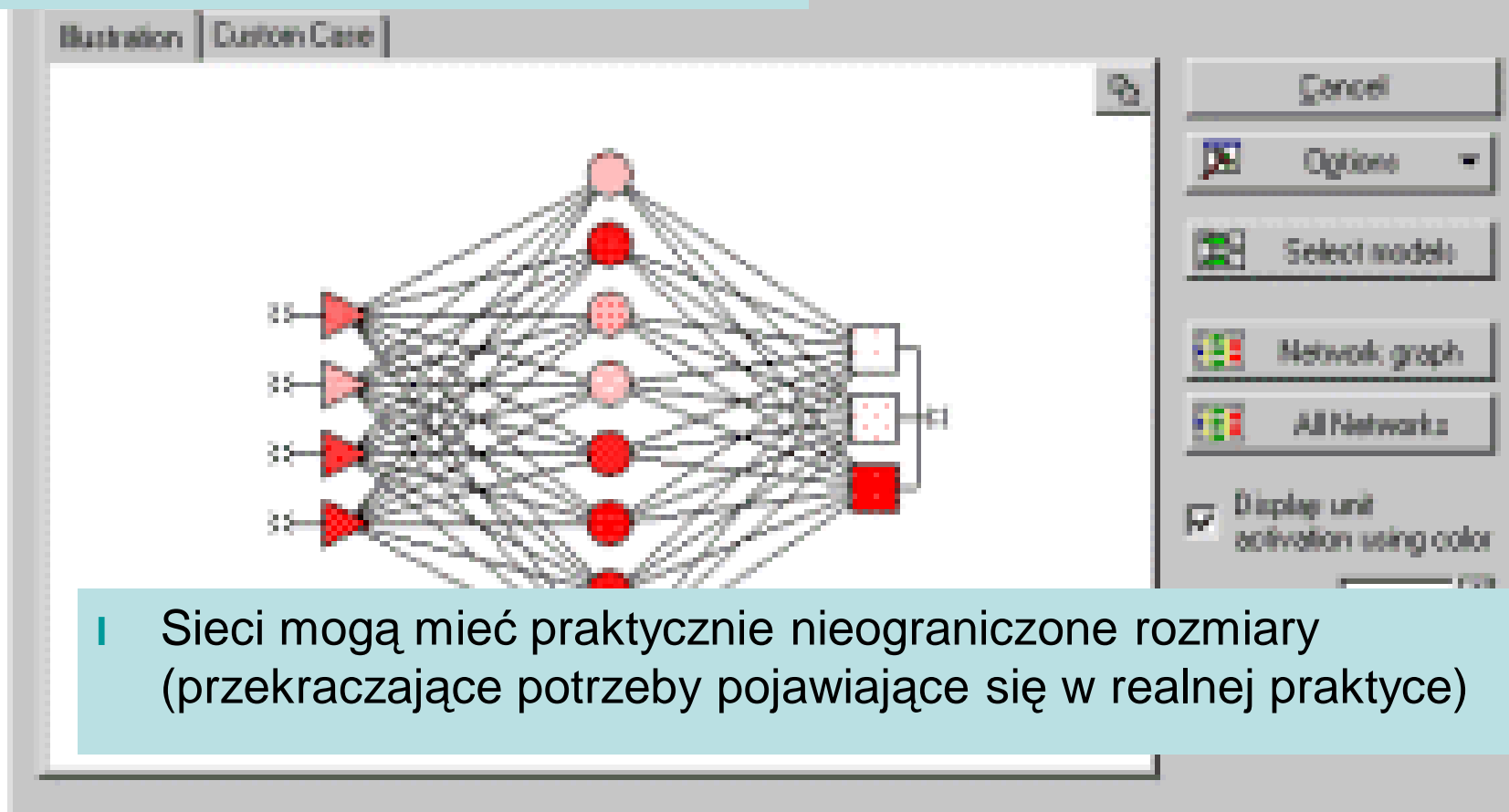
All Networks

Display unit activation using color

Input case: 133

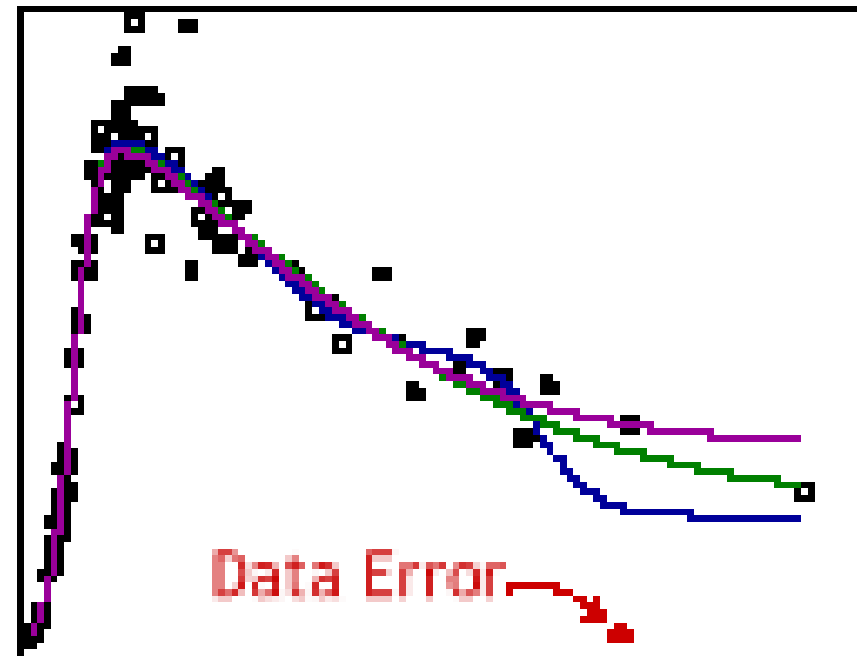
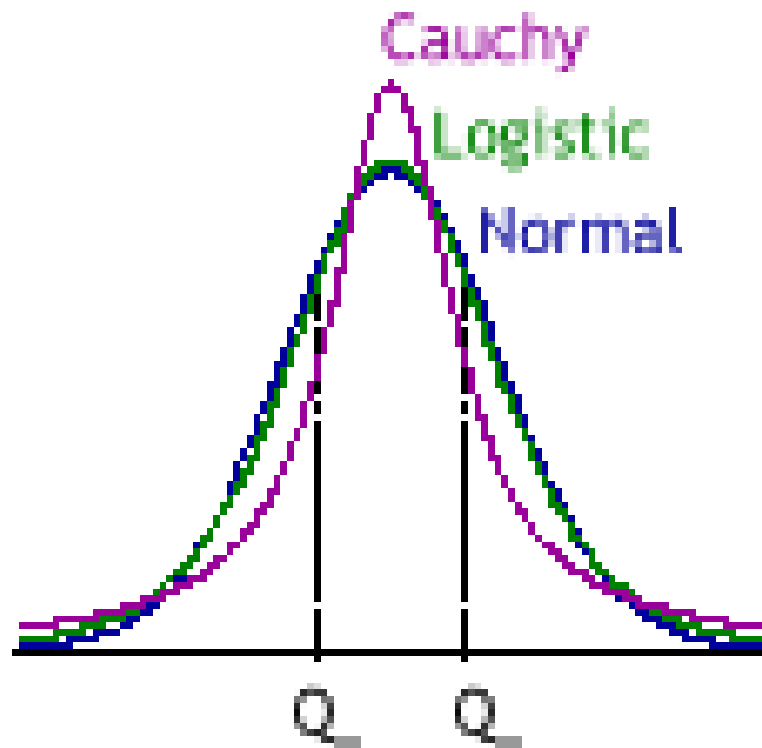
- I Mogą mieć do 128 warstw (w połączonych sieciach) i dowolną liczbę neuronów.
- I Ograniczenia programu są związane wyłącznie z zasobami sprzętowymi komputera.

	Test Part.	Train Error
	0.918919	0.201139
	0.691932	0.248233
	0.691932	0.176084
	0.000000	0.159535
	0.780794	0.180450
	0.780794	0.275002
	0.000000	0.000000
	0.780794	0.147924



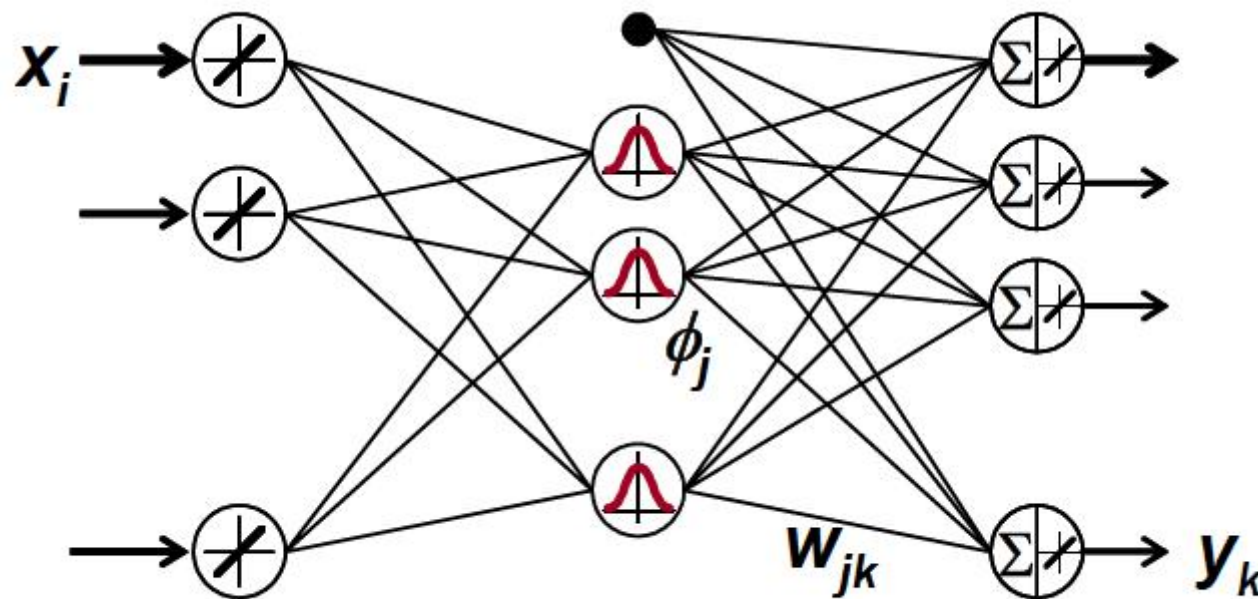
- I Sieci mogą mieć praktycznie nieograniczone rozmiary (przekraczające potrzeby pojawiające się w realnej praktyce)

Sieci RBF bywają nadmiernie wrażliwe na nawet nieliczne błędy w danych uczących



Sieć RBF

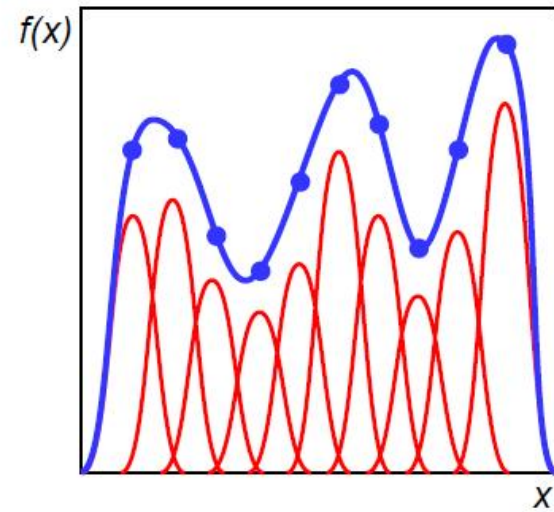
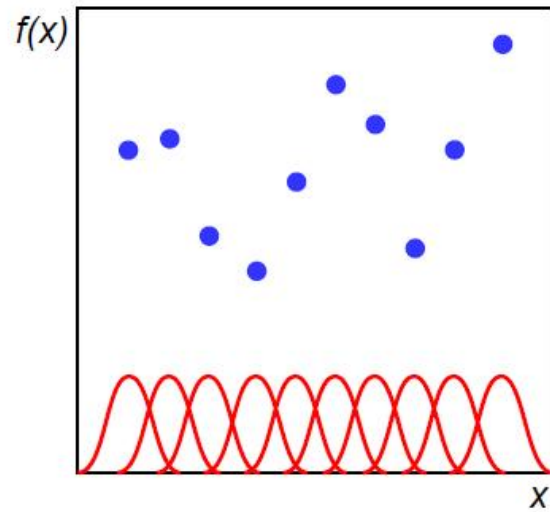
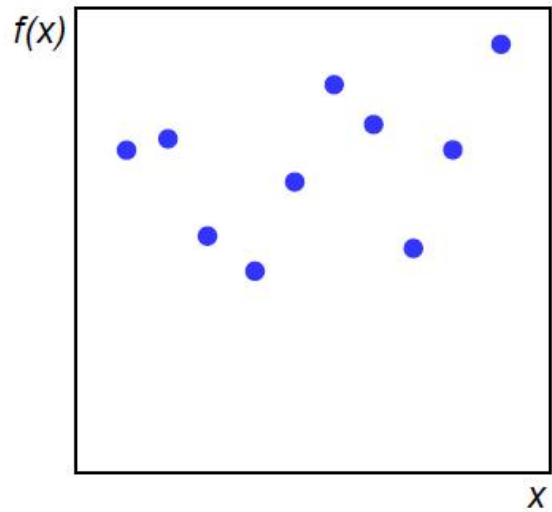
Ma ona strukturę dwuwarstwową, warstwa ukryta realizuje odwzorowanie nieliniowe realizowane przez neurony radialnej funkcji bazowej. Neuron wyjściowy jest liniowy, a jego rolą jest sumowanie wagowe sygnałów pochodzących od neuronów warstwy ukrytej.



Ogólna postać
sieci radialnej RBF

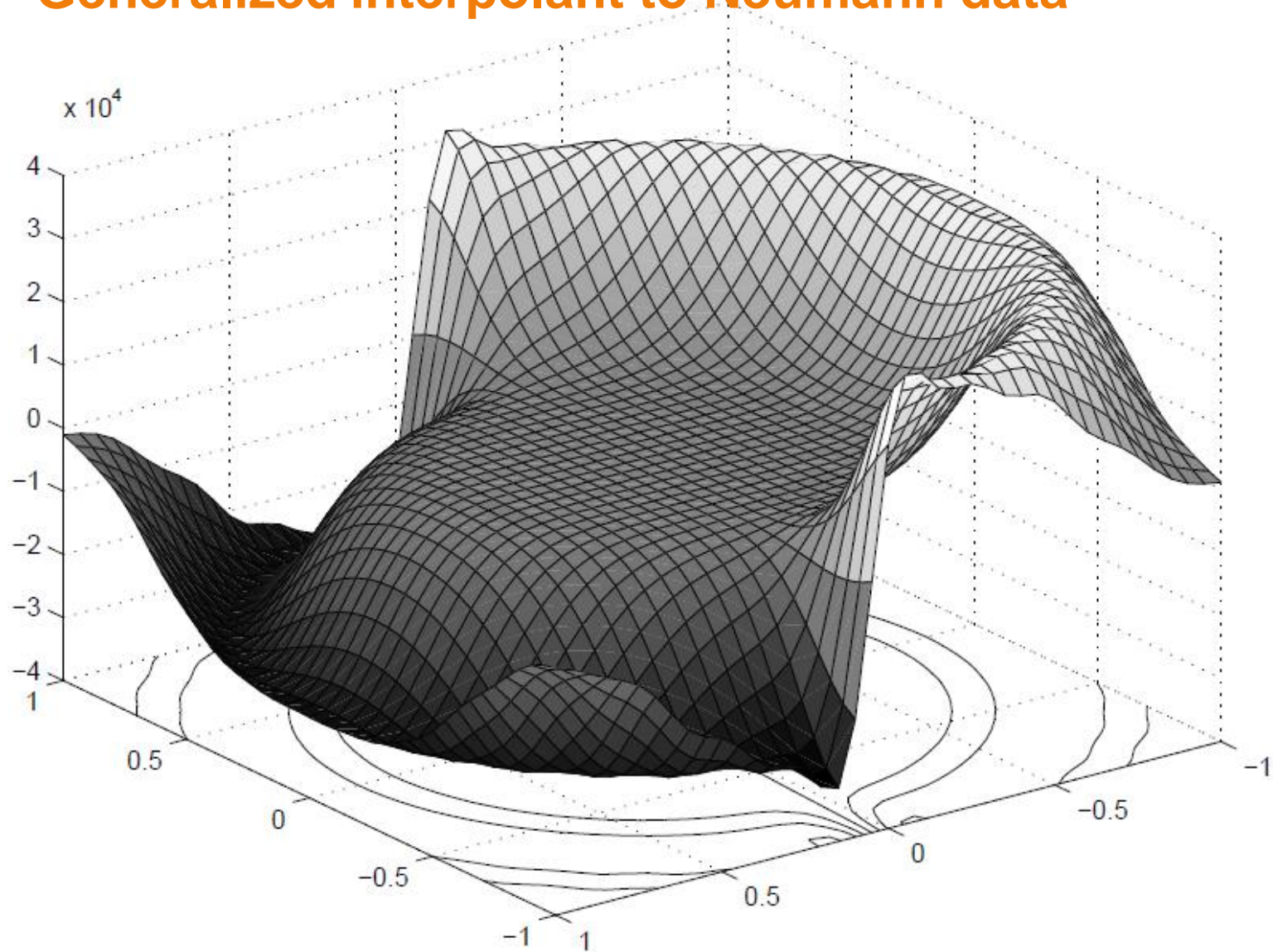


Introduction to Pattern Analysis
Ricardo Gutierrez-Osuna
Texas A&M University



Introduction to Pattern Analysis
Ricardo Gutierrez-Osuna
Texas A&M University

Generalized interpolant to Neumann data



R. Schaback. "A Practical Guide to Radial Basis Functions., April 2007.

Sieć radialna a sieć sigmoidalna

Sieci neuronowe o radialnych funkcjach bazowych znalazły zastosowanie zarówno w rozwiązywaniu problemów klasyfikacyjnych, zadaniach aproksymacji funkcji wielu zmiennych, jak i zagadnieniach predykcji..... tych obszarach zastosowań gdzie funkcje sigmoidalne mają ugruntowaną pozycję.

W stosunku do sieci wielowarstwowych o sigmoidalnych funkcjach aktywacji wyróżniają się pewnymi właściwościami szczególnymi, umożliwiającymi lepsze odwzorowanie cech charakterystycznych modelowanego procesu.

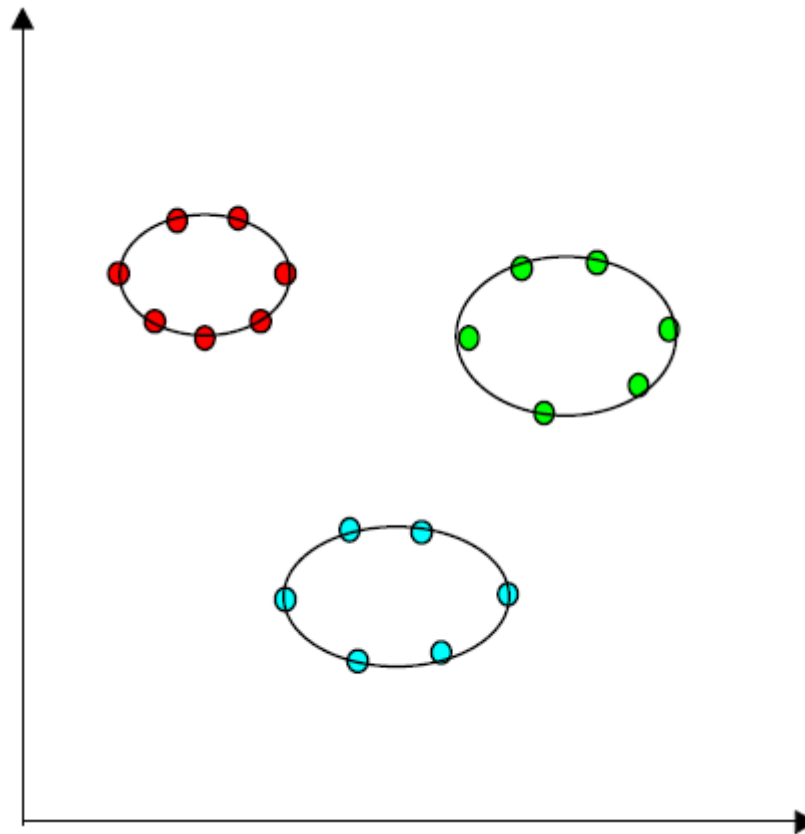
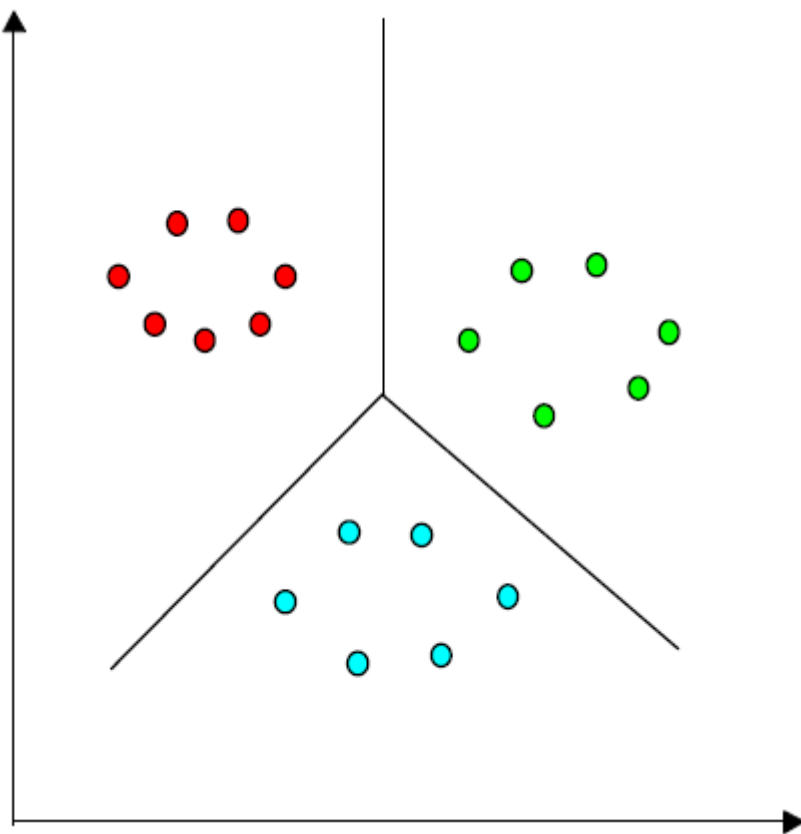
Bazowe funkcje radialne

Sieci typu radialnego stanowią naturalne uzupełnienie sieci sigmoidalnych.

→ Neuron sigmoidalny reprezentował w przestrzeni wielowymiarowej *hiperplaszczynę* separującą tę przestrzeń na dwie kategorie (klasy), w których był spełniony odpowiedni warunek, albo $\sum W_{ij}x_j > 0$ albo $\sum W_{ij}x_j < 0$.

→ Neuron radialny z kolei reprezentuje *hipersferę*, dokonującą podziału kołowego wokół punktu centralnego.

Sieć radialna a sieć sigmoidalna



Sieć radialna a sieć sigmoidalna

Sieć sigmoidalna:

Działanie funkcji rozciąga się od określonego punktu w przestrzeni aż do nieskończoności, reprezentuje aproksymację globalną funkcji zadanej. Nie ma niemożności fizycznego powiązania obszaru aktywności neuronu z odpowiednim obszarem danych uczących, trudności z określeniem optymalnego punktu startowego z procesie uczenia.

Sieć radialna:

Bazuje na funkcjach mających wartość niezerową jedynie w określonej przestrzeni tylko wokół centrów, realizuje aproksymację typu lokalnego, której zasięg działania jest bardzo ograniczony. Można się spodziewać że zdolności do uogólniania są gorsze niż dla sieci sigmoidalnych. Łatwość powiązania parametrów funkcji bazowych z fizycznym rozmieszczeniem danych w obszarze parametrów. Łatwość uzyskania dobrych wartości startowych w procesie uczenia pod nadzorem.

Sieć radialna a sieć sigmoidalna

Sieć radialna:

Przestrzenie decyzyjne tworzone w sieciach radialnych są stosunkowo proste i w sposób naturalny kształtowane. Sieć dostarcza nie tylko informacji do jakiej klasy należy wzorzec testujący, ale wskazuje również na ewentualną możliwość utworzenia oddzielnej klasy.

Na ogół uważa się, że sieci radialne lepiej niż sieci sigmoidalne nadają się do takich zadań klasyfikacyjnych jak wykrywanie uszkodzeń w różnego rodzaju systemach, rozpoznawanie wzorców, itp.

Znaczną zaletą sieci radialnych jest znacznie uproszczony algorytm uczenia. Przy istnieniu tylko jednej warstwy ukrytej i *ściśłym powiązaniu aktywności neuronu z odpowiednim obszarem przestrzeni danych uczących*, punkt startowy uczenia jest znacznie bliżej rozwiązania optymalnego, niż jest to możliwe w sieciach wielowarstwowych.

Sieć radialna a sieć sigmoidalna

Sieć radialna cont.:

Dodatkowo, możliwe jest odseparowanie etapu doboru parametrów funkcji bazowych od doboru wartości wag sieci (algorytm hybrydowy), co może przyspieszyć i uprościć proces uczenia. Przy zastosowaniu ortogonalizacji proces optymalnego kształtowania struktury sieci jest stałym fragmentem uczenia, nie wymagającym żadnego dodatkowego wysiłku.

Liczba neuronów ukrytych decyduje w dużym stopniu o dokładności odwzorowania i zdolnościach uogólniania sieci. W przypadku sieci radialnej problem doboru liczby neuronów ukrytych jest o wiele prostszy niż w sieciach sigmoidalnych, ze względu na lokalny charakter aproksymacji reprezentowany przez poszczególne funkcje bazowe.

Sieć RBF może być na różne sposoby hybrydyzowana - może być na przykład uczona algorytmem Kohonena i LVQ, co jest alternatywą do przypisywania centrów odzwierciedlającego rozkład danych. Warstwa wyjściowa (liniowa lub nie) może być uczona którymkolwiek z iteracyjnych algorytmów dla warstw z iloczynem skalarnym.

Dobór parametrów funkcji radialnych

Znacznie lepsze rezultaty można uzyskać przez zastosowanie

samoorganizującego się procesu podziału danych uczących

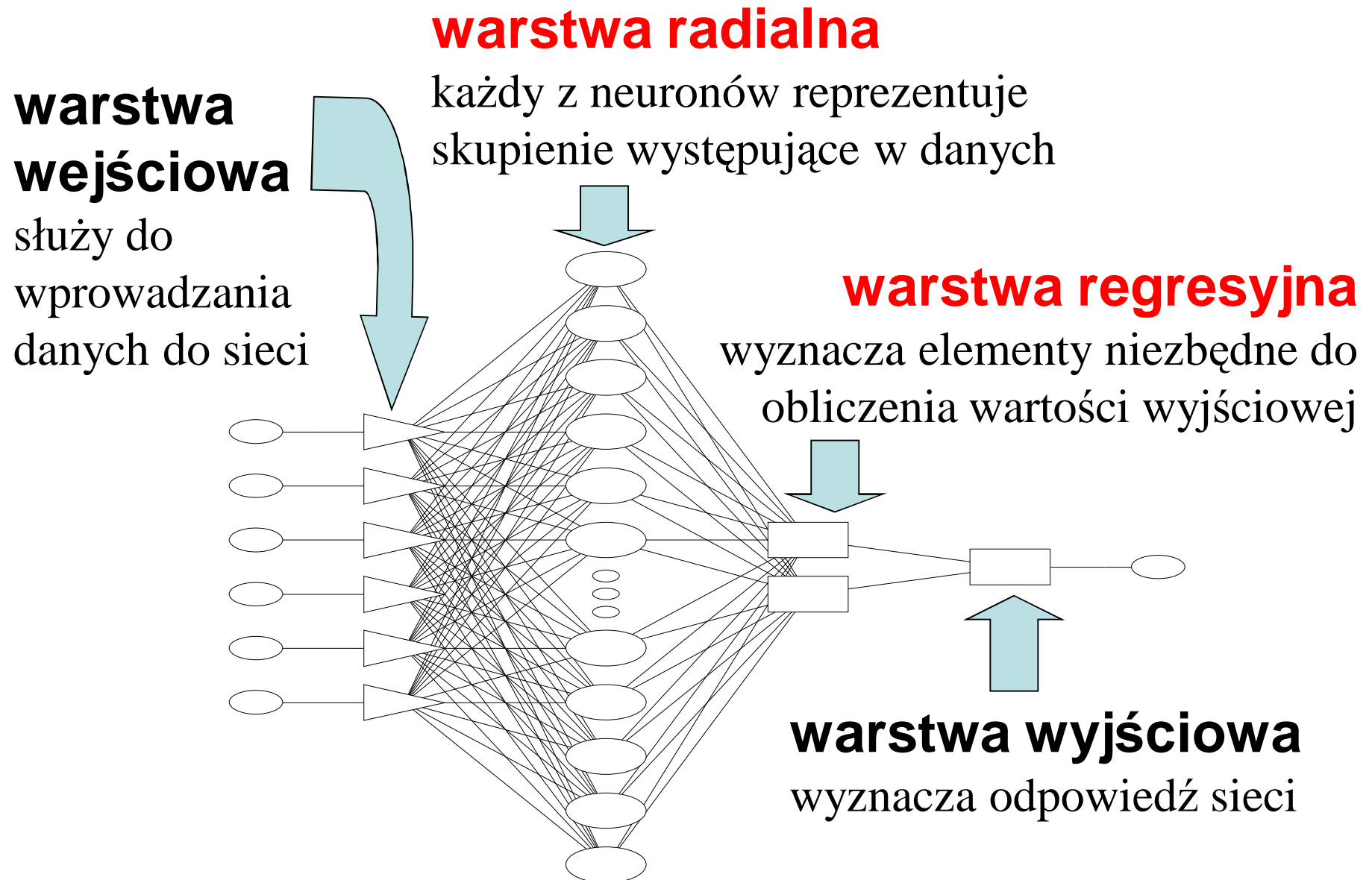
na klastry w jednej z jego licznych odmian.

→ Centrum klastra jest utożsamiane z centrum odpowiedniej funkcji radialnej.

→ Liczba tych funkcji jest równa liczbie klastrów i może być korygowana przez algorytm samoorganizacji.

→ Proces podziału danych na klastry może być przeprowadzany metodą K-uśrednień. Aparat matematyczny zaangażowany w tą procedurę jest dość skomplikowany....

A tak wygląda struktura innej praktycznie użytecznej sieci klasy GRNN



Idea działania sieci realizujących regresję uogólnioną

(GRNN - *Generalized Regression Neural Network*)

- Wejściowe wektory uczące dzielone są na **skupienia** - w szczególnym przypadku każdy wektor tworzy oddzielne skupienie,
- Dla każdego skupienia znana jest wartość zmiennej objaśnianej (wyjście sieci),
- wartość zmiennej objaśnianej dla dowolnego wektora wejściowego szacowana jest jako średnia ważona liczona z wartości tej zmiennej dla skupień - wagi uzależnione są od odległości wejścia od centrów skupień.