Podstawy Sztucznej Inteligencji
Sztuczne Sieci Neuronowe

Krzysztof Regulski,
WIMiIP, KISiM,
regulski@agh.edu.pl
B5, pok. 408

Inteligencja

- Czy inteligencja jest jakąś jedną dziedziną, czy też jest to nazwa dla zbioru odrębnych i niepowiązanych zdolności?
- Co zyskujemy w procesie uczenia się?
- Co to jest intuicja?
- Czy inteligencja może być nabyta wskutek nauki lub obserwacji, czy też jest jakoś uwarunkowana wewnętrznie?
- Jak wiedza wpływa na wzrost inteligencji?
- Czy inteligencja to szczegółowa wiedza o jakiejś dziedzinie, czy zbiór związanych ze sobą różnych zdolności?

Inteligencja

Inteligencja jest zdolnością do sprawnego rozwiązywania zadań intelektualnych, które zazwyczaj uchodzą za trudne.

… są trudne tak długo, jak długo nie są znane algorytmy ich rozwiązywania, potem przestają być traktowane jako zadania sztucznej inteligencji

w ten sposób sztuczna inteligencja nigdy nie ma żadnych osiągnięć

Czy to jest trudny problem ?

$$9873126986841431698425168435 1 \times 9853168463159684631986435 41684$$

A to:

”Kochanie, kup ładny kawałek wołowiny…”

Czy nam to szybko grozi?

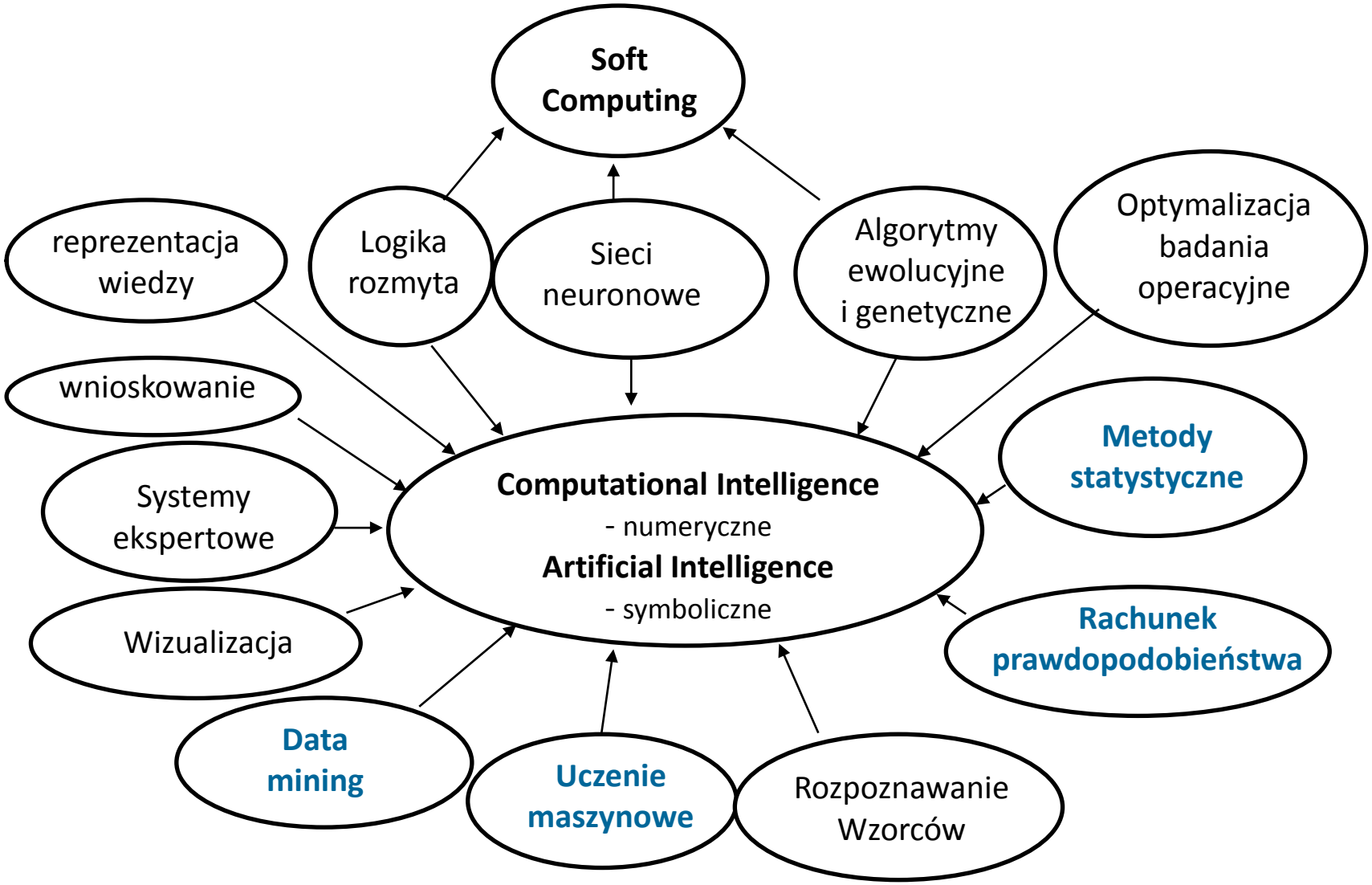
Robot kolejkowy EWA-1

-Pan tu nie stał, pan nie jest w ciąży.

Krzysztof Manc
(Wynalazca)

- Moja konstrukcja jest optymalna, tylko ludzie nie dorośli do tego. Wolą sami stać w kolejkach.

Zagadnienia Sztucznej Inteligencji (AI)



Przykłady zadań sztucznej inteligencji

- dokonywanie ekspertyz ekonomicznych, prawnych, technicznych, medycznych (ocena)
- wspomaganie podejmowania decyzji (doradzanie)
- rozpoznawanie obrazów, twarzy, wzorców, etc.
- optymalizacja (harmonogramowanie, alokacja zasobów, planowanie tras)
- generacja nowej wiedzy (poszukiwanie zalezności, tendencji, reguł, etc – data mining)
- prognozowanie zjawisk ekonomicznych, przyrodniczych
- rozumienie języka naturalnego
- sterowanie urządzeniami (roboty etc)
- i inne…

Zadania sztucznej inteligencji

Wnioskowanie

» procesem stosowania reguł wnioskowania w sposób skutecznie i efektywnie prowadzący do określonego celu wnioskowania, którym zazwyczaj jest uzyskania pewnego docelowego stwierdzenia.

Uczenie się

» proces zmiany zachodzącej w systemie na podstawie doświadczeń, która prowadzi do poprawy jego jakości działania rozumianej jako sprawność rozwiązywania stojących przed systemem zadań.

Przeszukiwanie

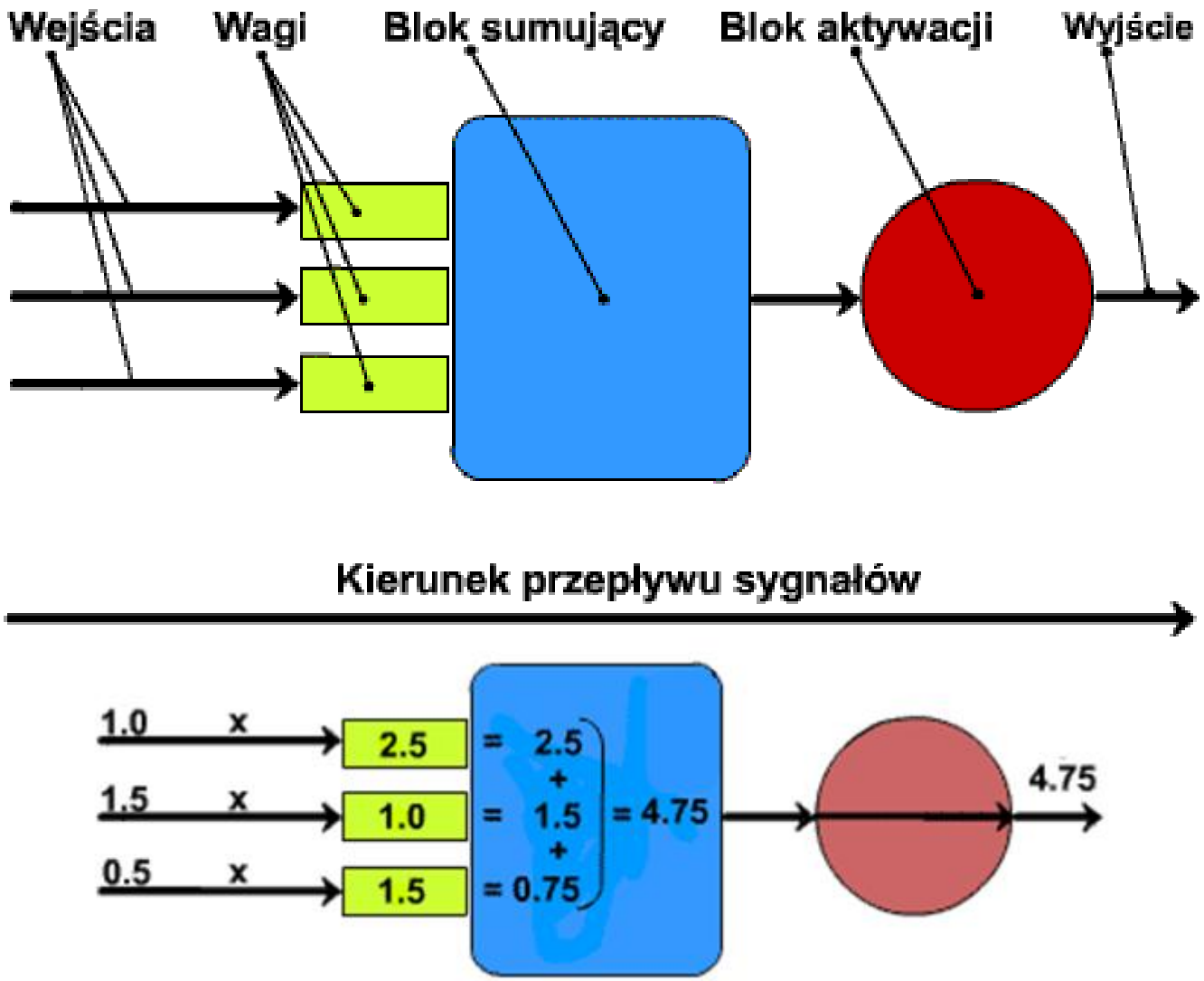
» Znajdowanie zadawalających rozwiązań bez pełnego przeglądania wszystkich możliwości, a więc jak najbardziej efektywne przeszukiwanie przestrzeni rozwiązań

» Przeszukiwanie przestrzeni jest procesem stosowania dostępnych operatorów do przemieszczania się między różnymi stanami w tej przestrzeni, tak aby ostatecznie trafić ze stanu początkowego do jednego ze stanów końcowych

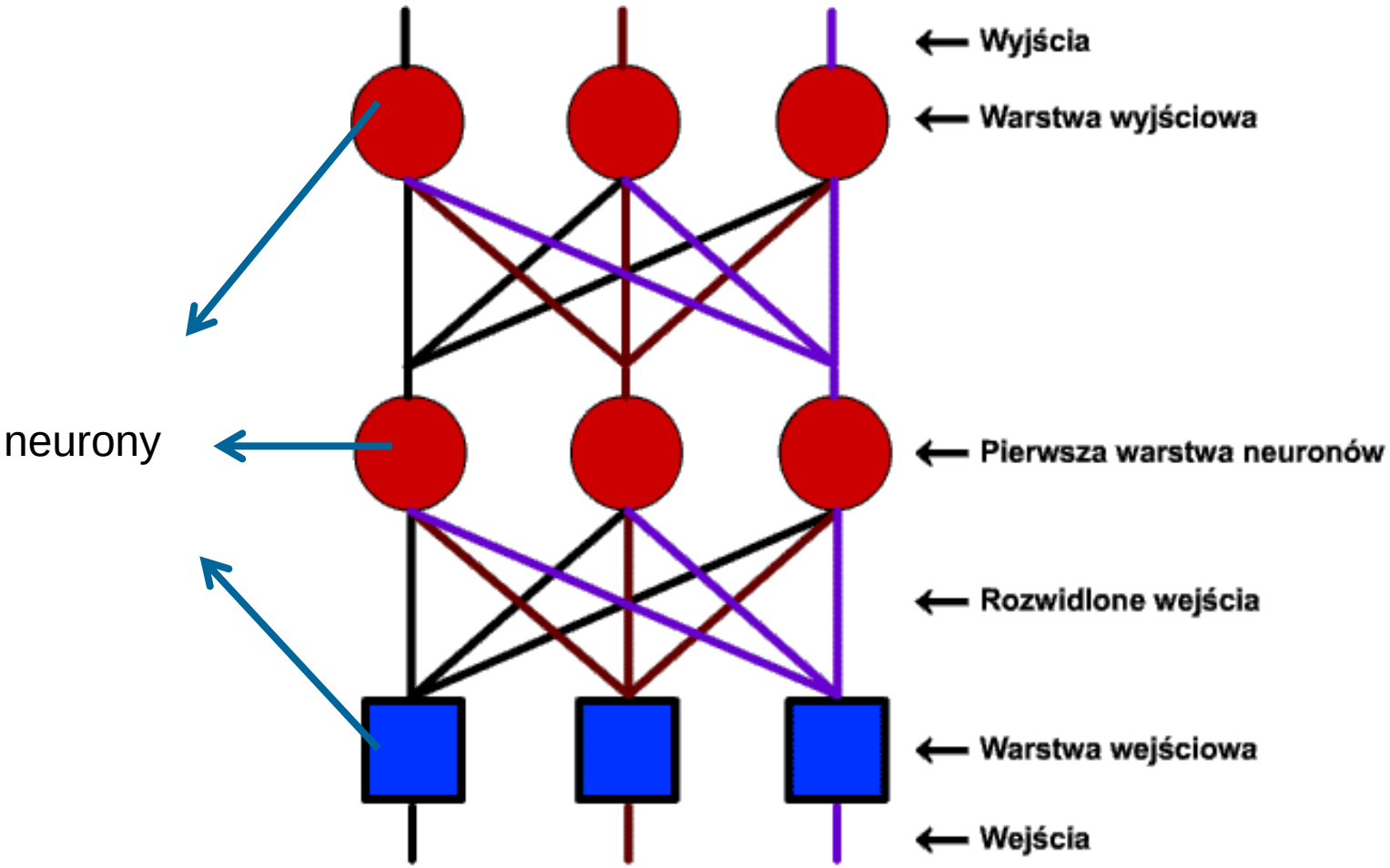
Uczenie się –

Sztuczne Sieci Neuronowe

Sztuczny neuron, McCulloch i Pitts,1943



Z czego składa się sztuczna sieć neuronowa?



Jak to wszystko działa?

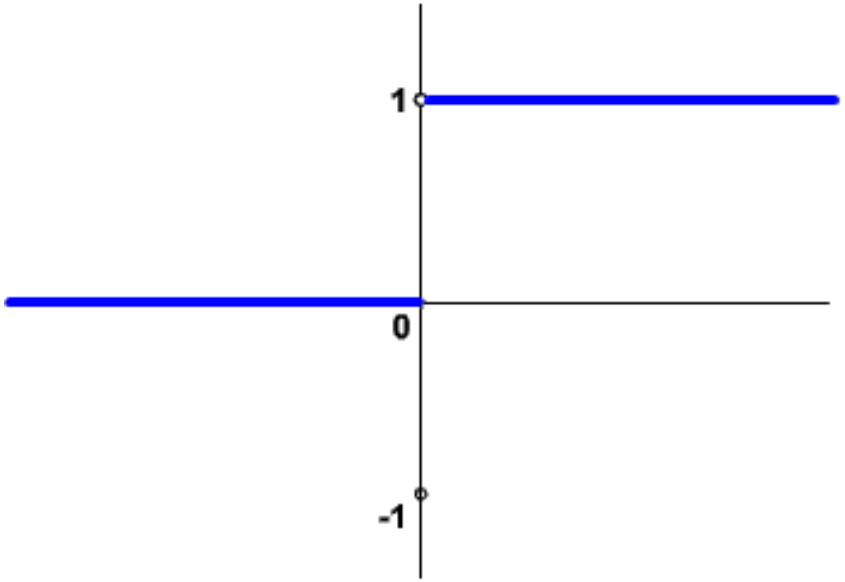
- Rozproszone są:
  - program działania
  - baza wiedzy
  - dane
  - proces obliczania
- Sieć działa zawsze jako całość, równolegle
- Wszystkie elementy mają wkład we wszystkie czynności
- Duża zdolność do działania nawet po uszkodzeniach
- **wagi** – pojęcie kluczowe

Po co są wagi?

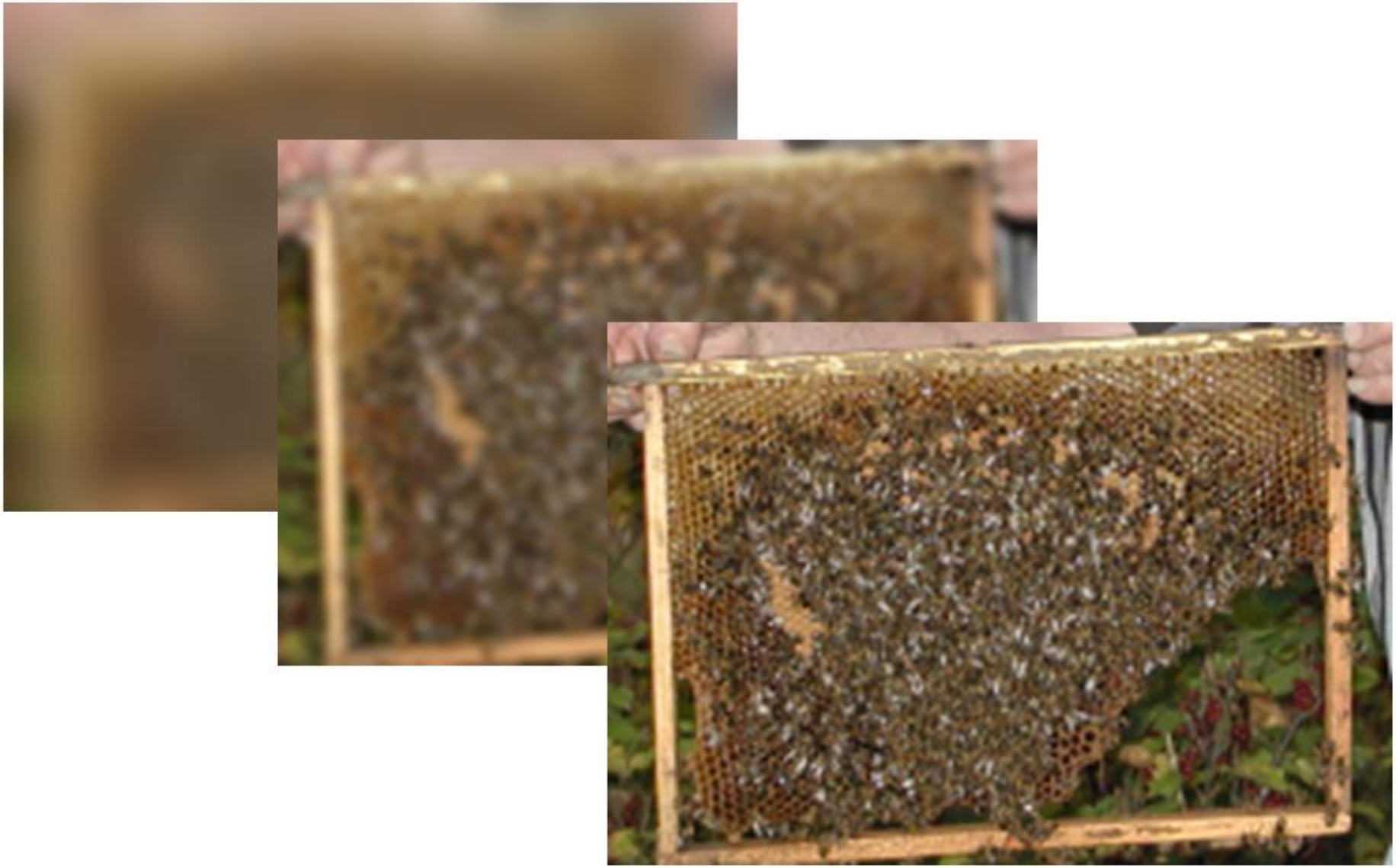
WAGA "NOWY"	WAGA „ŁADNY"	**Co na to neuron?**
5	1	Lubi nowe samochody; ich wygląd mało się liczy.
1	-1	Wszystko mu jedno...
-5	4	Samochód musi być ładny, ale stary.
4	5	Lubi nowe i piękne auta.
-4	-4	Ekstrawagancki neuron - lubi stare, brzydkie maszyny.

Po co są funkcje aktywacji?

Wartość sygnału wyjściowego zależy od rodzaju funkcji w bloku aktywacji

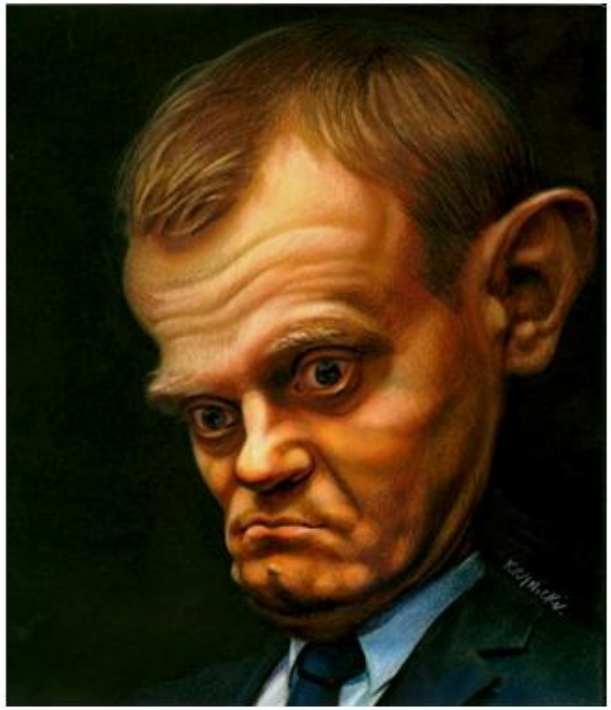


Rozpoznawanie



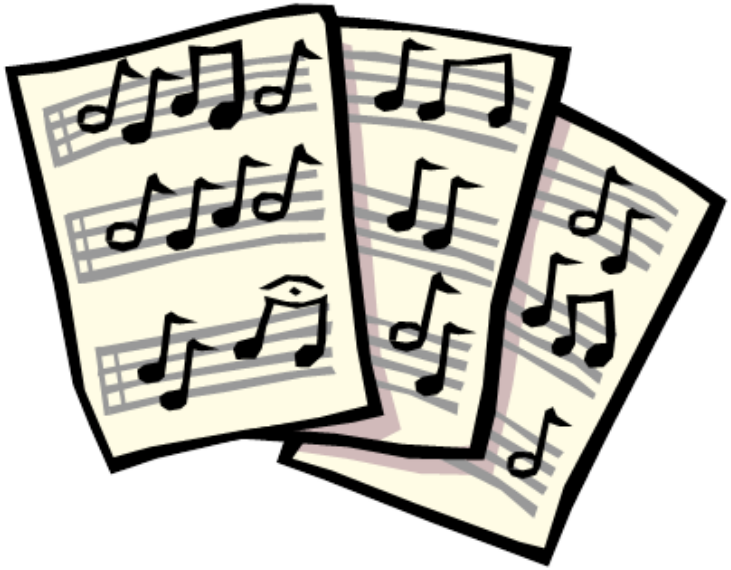
Rozpoznawanie

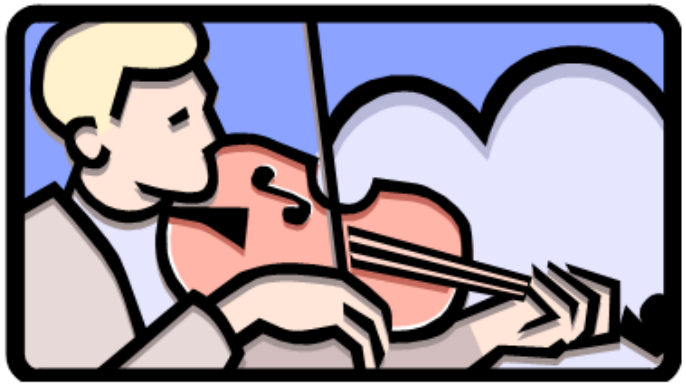
W jaki sposób błyskawicznie rozpoznajemy twarze znanych osób? (nawet zniekształcone)



Cudów nie ma! Rozpoznajemy tylko taką twarz, której nasz mózg się nauczył.

Kojarzenie





Przewidywanie

Do czego służy sieć neuronowa?

... do wszelkiego rodzaju

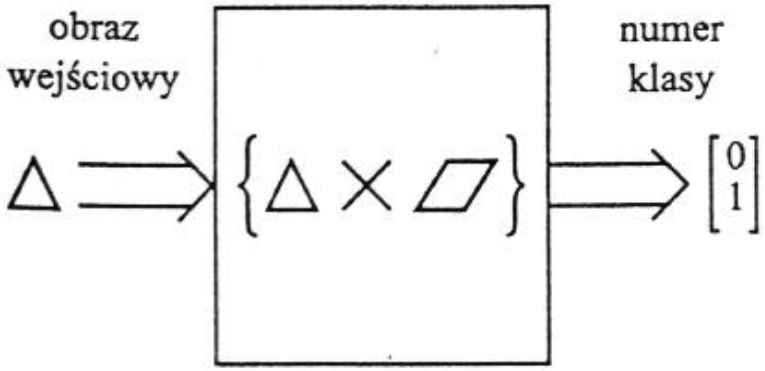
- rozpoznawania,
- kojarzenia,
- przewidywania

... praktycznie we wszystkich dziedzinach życia, nauki, techniki, przemysłu, ...

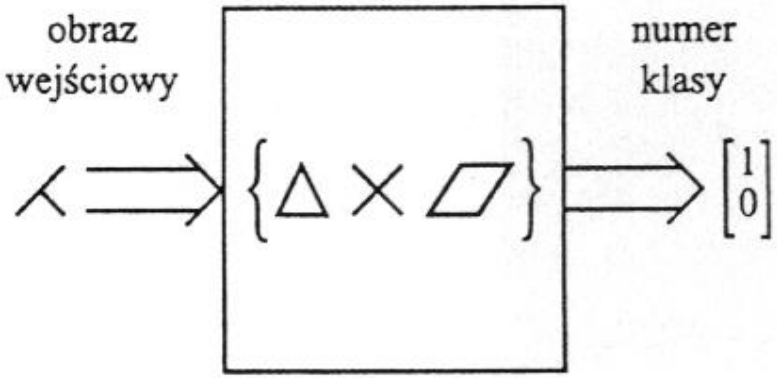
Klasyfikacja

Klasyfikacja. Do budowy klasyfikatorów wykorzystuje się neurony o wyjściach binarnych.
Przy prezentacji obrazu uaktywnia się tylko jeden element wyjściowy.
Obraz wejściowy zostaje przyporządkowany określonej klasie reprezentowanej przez aktywny element.

- klasyfikacja



- rozpoznawanie



Zadaniem procesu uczenia jest podział wzorców uczących na klasy obrazów zbliżonych do siebie i przyporządkowanie każdej klasie osobnego elementu wyjściowego

Co ma wpływ na sygnały wyjściowe?

Sygnał wyjściowy neuronu zależy od:

- sygnałów wejściowych,
- współczynników wagowych,
- funkcji aktywacji,
- budowy sieci

Uczenie

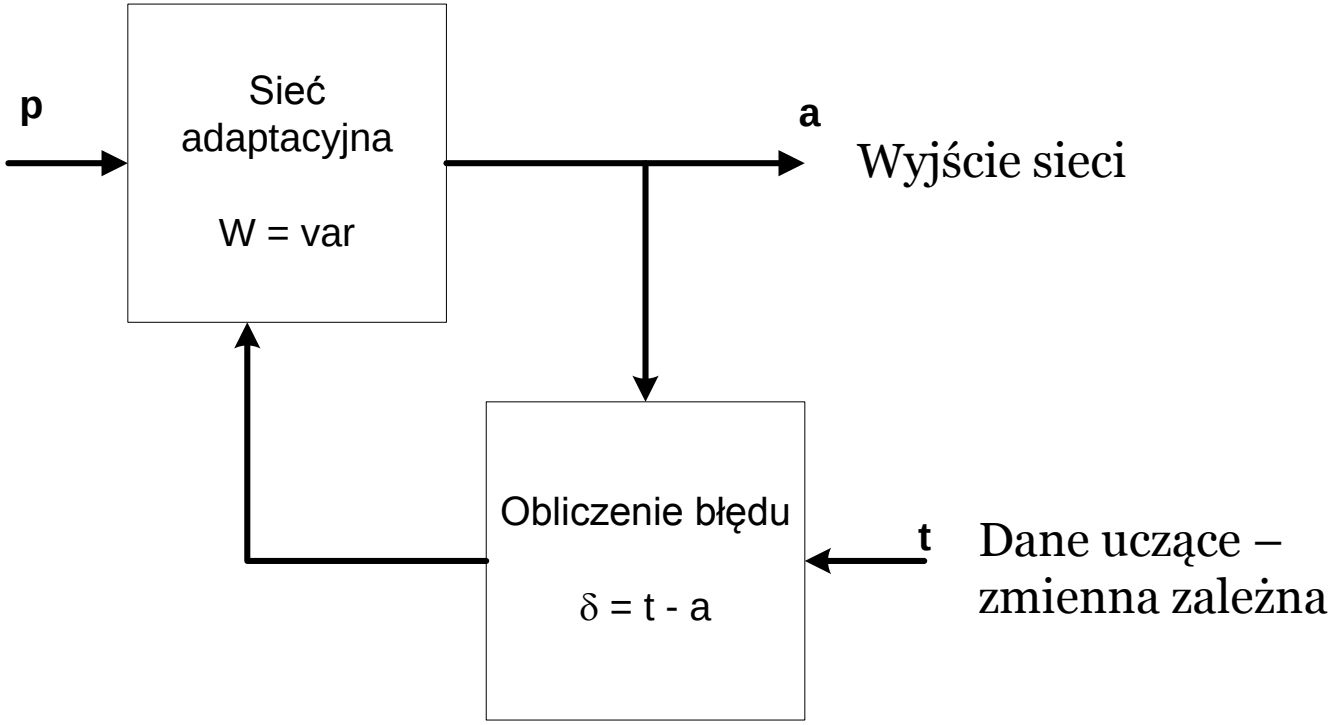
- Zmieniając wartości współczynników wagowych możemy zmieniać zależności pomiędzy wektorami sygnałów wejściowych i wyjściowych sieci.

- Jak ustawić te współczynniki aby uzyskać żądane zależności pomiędzy tymi wektorami?

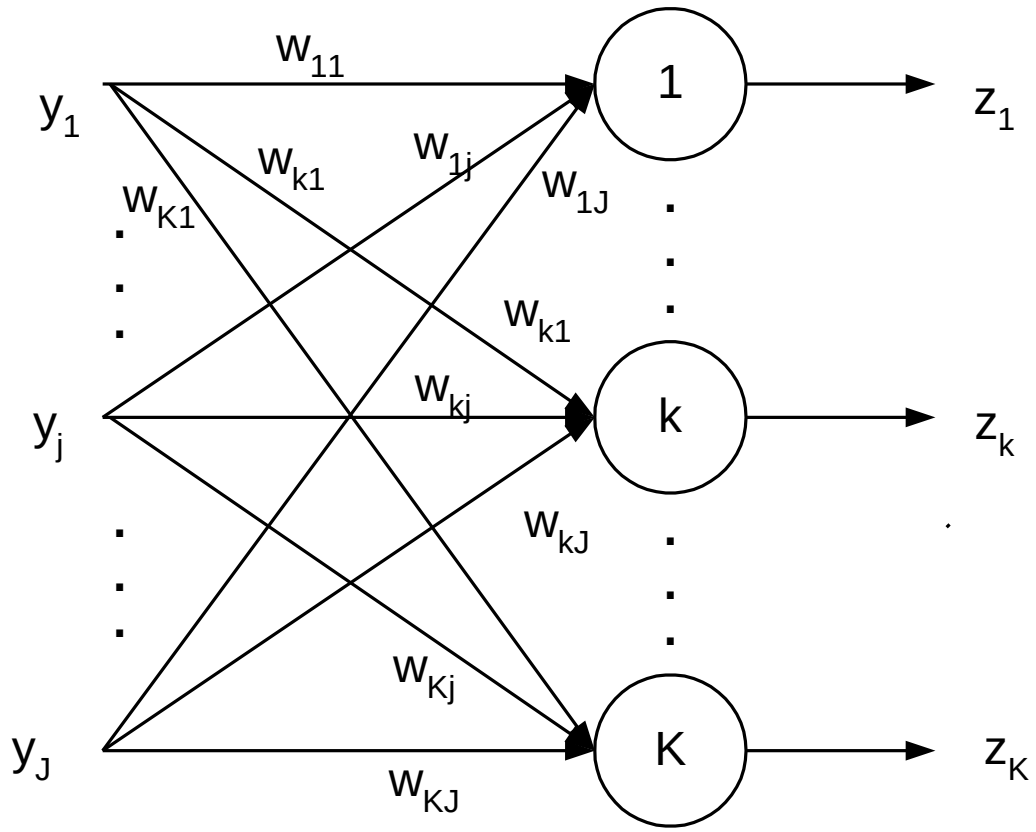
- Sieci neuronowych się nie programuje. Oznacza to, że wartości współczynników wag nie ustala się bezpośrednio, ulegają one zmianom podczas procesu nazywanego uczeniem sieci.

Uczenie nadzorowane

- W procesie uczenia z nauczycielem, dla nowych wartości sygnałów wejściowych nauczyciel (zmienna zależna) podpowiada pożądaną odpowiedź.
- *Błąd odpowiedzi służy do korygowania wag sieci*.



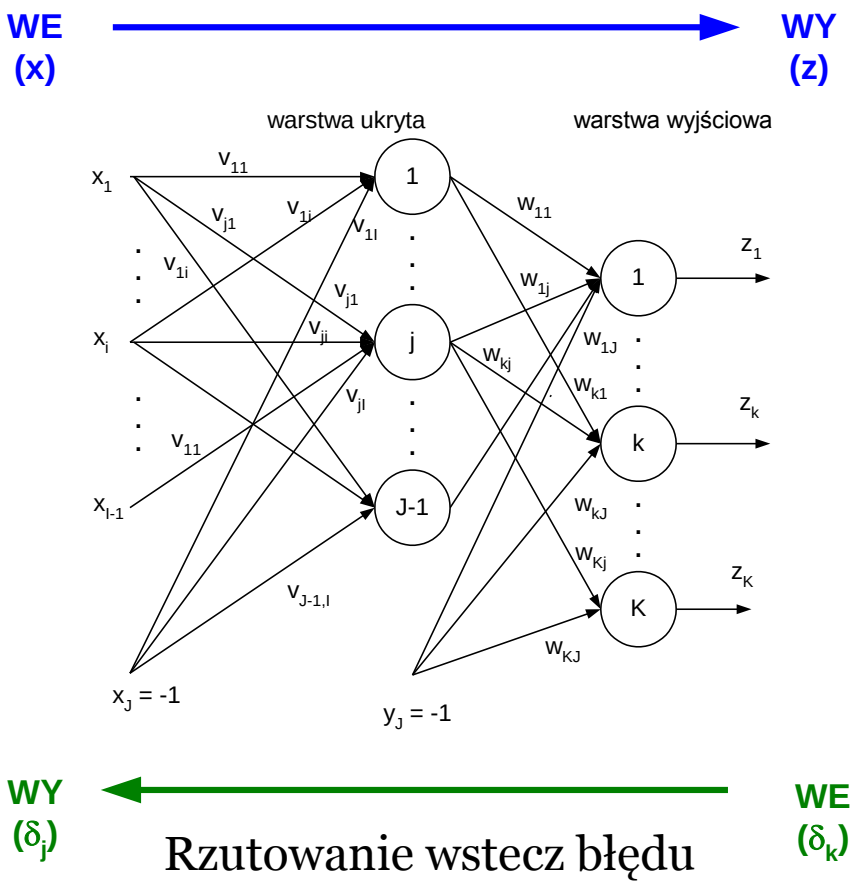
Uczenie sieci jednowarstwowej



minimalizacja błędu uczenia

Dla jakich wartości współczynników wagowych, błąd uczenia przyjmuje najmniejszą (minimalną) wartość?

Uczenie sieci wielowarstwowej
Algorytm wstecznej propagacji błędu



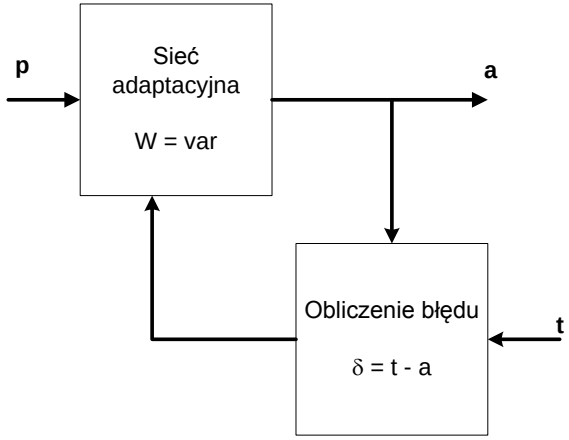
- Dla każdej próbnej liczby neuronów ukrytych tworzymy losowe wagi początkowe i uczymy sieć do chwili, gdy przestajemy osiągać poprawę.
- Sprawdzamy działanie sieci przy pomocy danych testowych.
- Jeżeli jakość sieci jest wyraźnie gorsza niż na zbiorze uczącym, to:

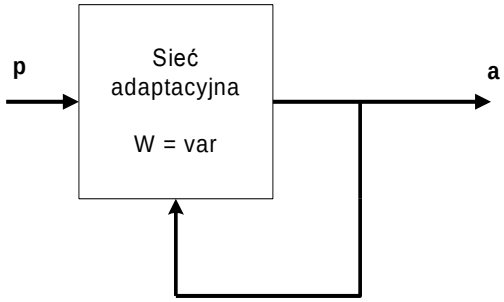
  - albo zbiór uczący jest zły (za mały lub mało reprezentatywny)
  - albo mamy za dużo neuronów ukrytych. Sieć jest przetrenowana.

Uczenie
bez nauczyciela

Uczenie bez nauczyciela

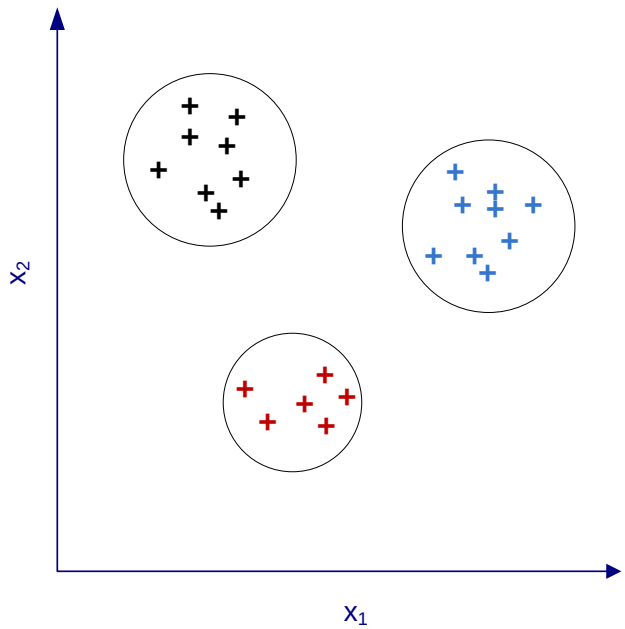
- W procesie uczenia z nauczycielem, dla nowych wartości sygnałów wejściowych nauczyciel podpowiada pożądaną odpowiedź.
- *Błąd odpowiedzi służy do korygowania wag sieci*.



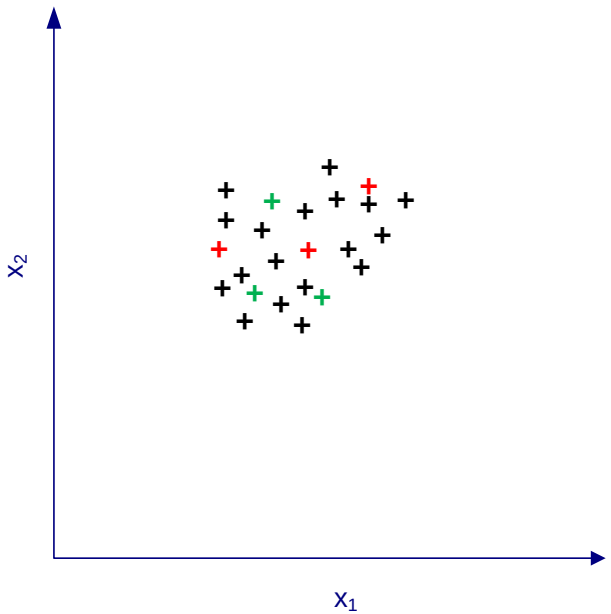


- Podczas uczenia bez nauczyciela pożądana odpowiedź nie jest znana.
- Sieć sama musi uczyć się poprzez analizę reakcji na pobudzenia (o których naturze wie mało albo nic).

Uczenie bez nauczyciela



Istnieją sytuacje, gdy sieć nie może się sama nauczyć, bez nauczyciela lub bez dodatkowej informacji o analizowanym problemie.



Uczenie „bez nauczyciela” jest możliwe, gdy mamy nadmiarowość danych wejściowych.

Reguła Hebba

Reguła Hebba (1949) oparta jest na doświadczeniu Pawłowa (odruch Pawłowa).

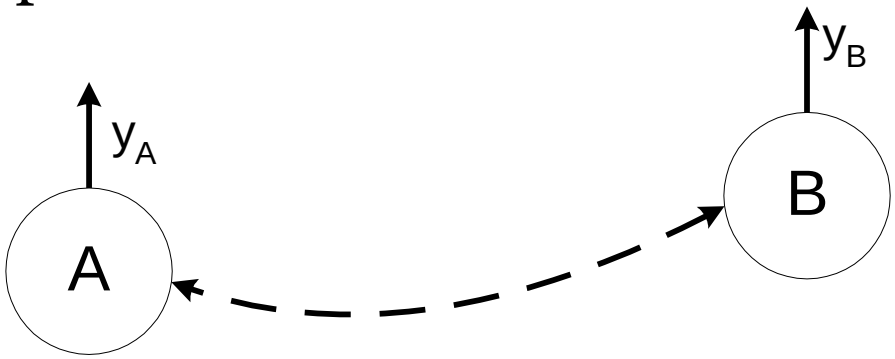
pokarm (BB)	ślinienie (OB)
pokarm (BB) + dzwonek (BW)	ślinienie (OW)
dzwonek (BW)	ślinienie (OW)

OB - odruch bezwarunkowy,
OW - odruch warunkowy,
BB - bodziec bezwarunkowy,
BW - bodziec warunkowy.

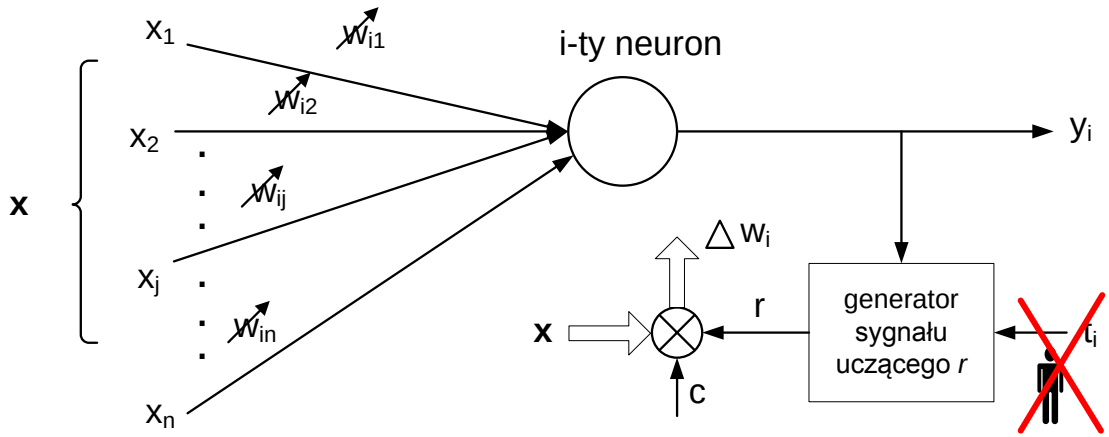
Reguła (model neuronu) Hebba (1949)

Często powtarzające się obrazy wejściowe dają silniejsza odpowiedź na wyjściu.

Jeżeli akson komórki A bierze systematycznie udział w pobudzaniu komórki B powodującym jej aktywację, to wywołuje to zmianę metaboliczną w jednej lub w obu komórkach, prowadzącą do wzrostu skuteczności pobudzania B przez A



Reguła Hebba



nazywa się także niekiedy uczeniem korelacyjnym

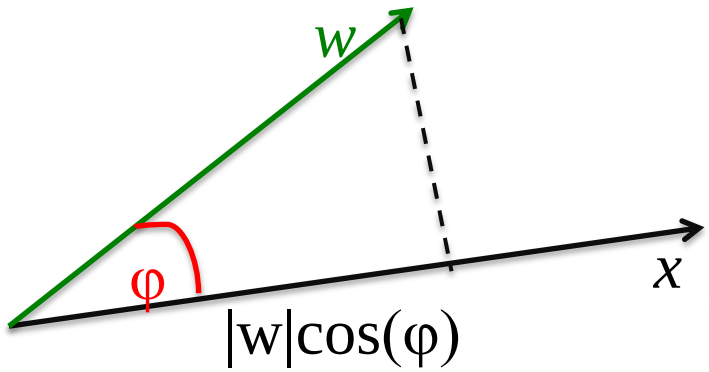
Sygnałem uczącym jest po prostu sygnał wyjściowy neuronu.

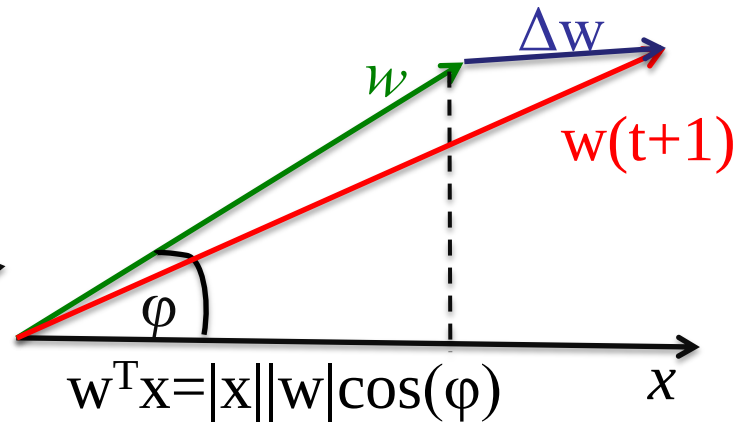
Korekta wektora wag:

$$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) + \Delta\mathbf{w}_i = \mathbf{w}_i(k) + cy_i\mathbf{x}$$

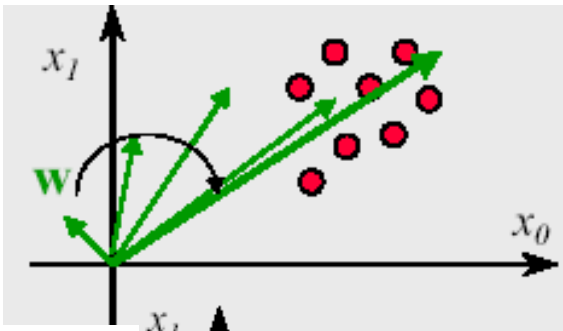
Często powtarzające się obrazy wejściowe dają więc silniejsza odpowiedź na wyjściu.

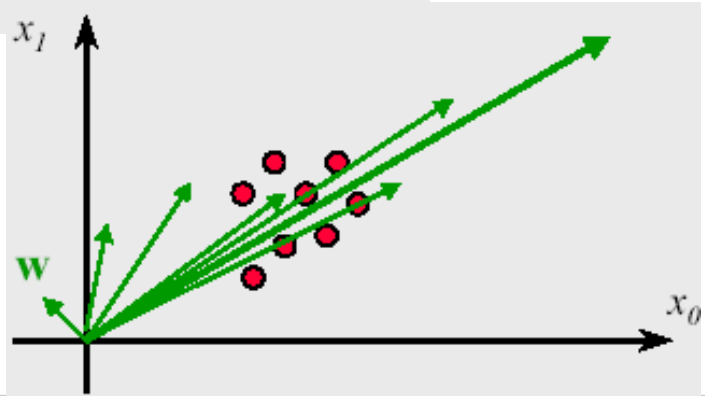
Interpretacja geometryczne reguły Hebba





Wektor wag przemieszcza się w kierunku środka ciężkości wielkości uczących





<u>Niebezpieczeństwo:</u>
Długość wektora wag rośnie w sposób nieograniczony
(o ile nie zapewnimy stopniowego zmniejszania współczynnika *c*)

Ile warstw ukrytych ?

Sieci o większej liczbie warstw ukrytych uczą się dłużej i proces uczenia jest słabiej zbieżny.

Rozbudowanie sieci do dwóch warstw ukrytych jest uzasadnione w zagadnieniach w których sieć ma nauczyć się funkcji nieciągłych.

Zaleca się rozpoczynać zawsze od jednej warstwy ukrytej.

Jeżeli pomimo zwiększenia liczby neuronów sieć uczy się źle, wówczas należy podjąć próbę dodania dodatkowej drugiej warstwy ukrytej i zmniejszyć jednocześnie liczbę neuronów ukrytych.

Ile neuronów w warstwach ukrytych ?

Również i tutaj brak jednoznacznej odpowiedzi

Zbyt mała liczba neuronów w warstwie ukrytej - mogą wystąpić trudności w uczenia sieci.

Zbyt duża liczba neuronów - wydłużenie czasu uczenia.

Może doprowadzić do tzw. nadmiernego dopasowania. Sieć uczy się wtedy "na pamięć" pewnych szczegółów i zatraca zdolność do uogólnień. Taka sieć uczy się wprost idealnie i wiernie odwzorowuje wartości zbioru uczącego. Jednakże wyniki testowania sieci na danych spoza zbioru uczącego są zazwyczaj bardzo słabe.

Jak konstruować sieć ?

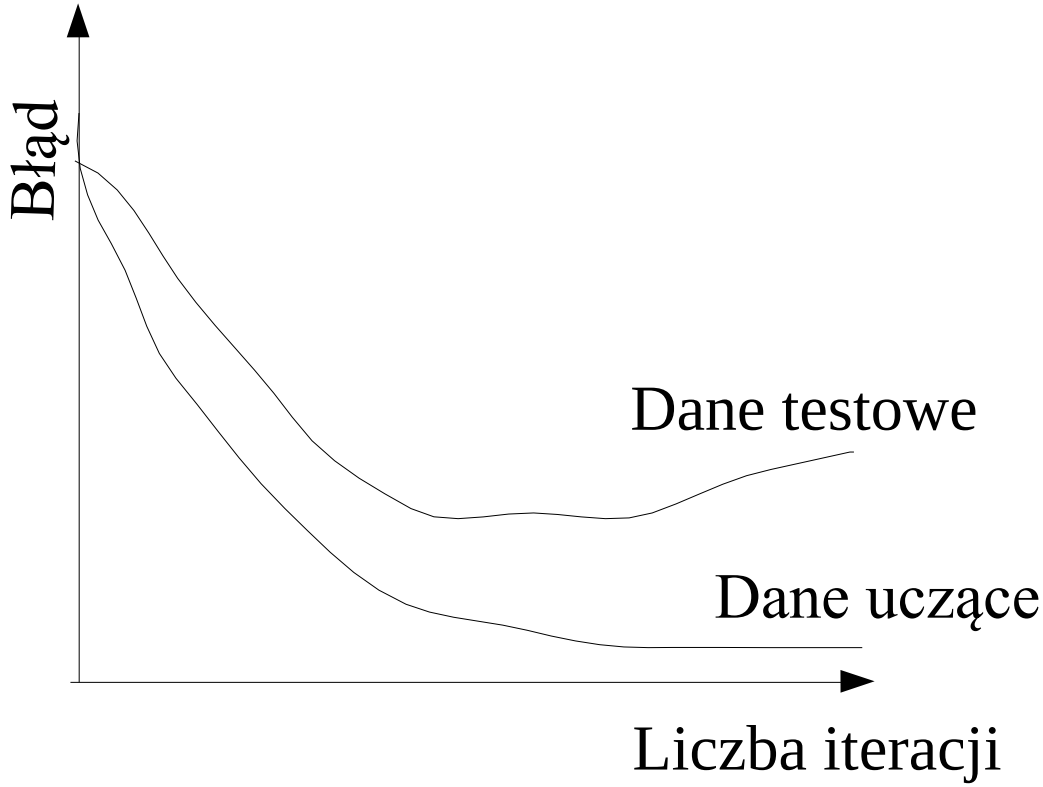
Reasumując:

Ze względu na fakt, że brak jest teoretycznych przesłanek do projektowania topologii sieci, zgodnie z zaleceniami podanymi przez Masters'a, we wstępnych pracach nad siecią dobrze jest przestrzegać następujących 3 zasad:

1. należy stosować tylko jedną warstwę ukrytą,
2. należy używać niewielu neuronów ukrytych,
3. należy uczyć sieć do granic możliwości.

Jak długo uczyć sieć ?

Oznaką przetrenowania jest wzrost błędu dla danych testowych, przy malejącym błędzie danych uczących



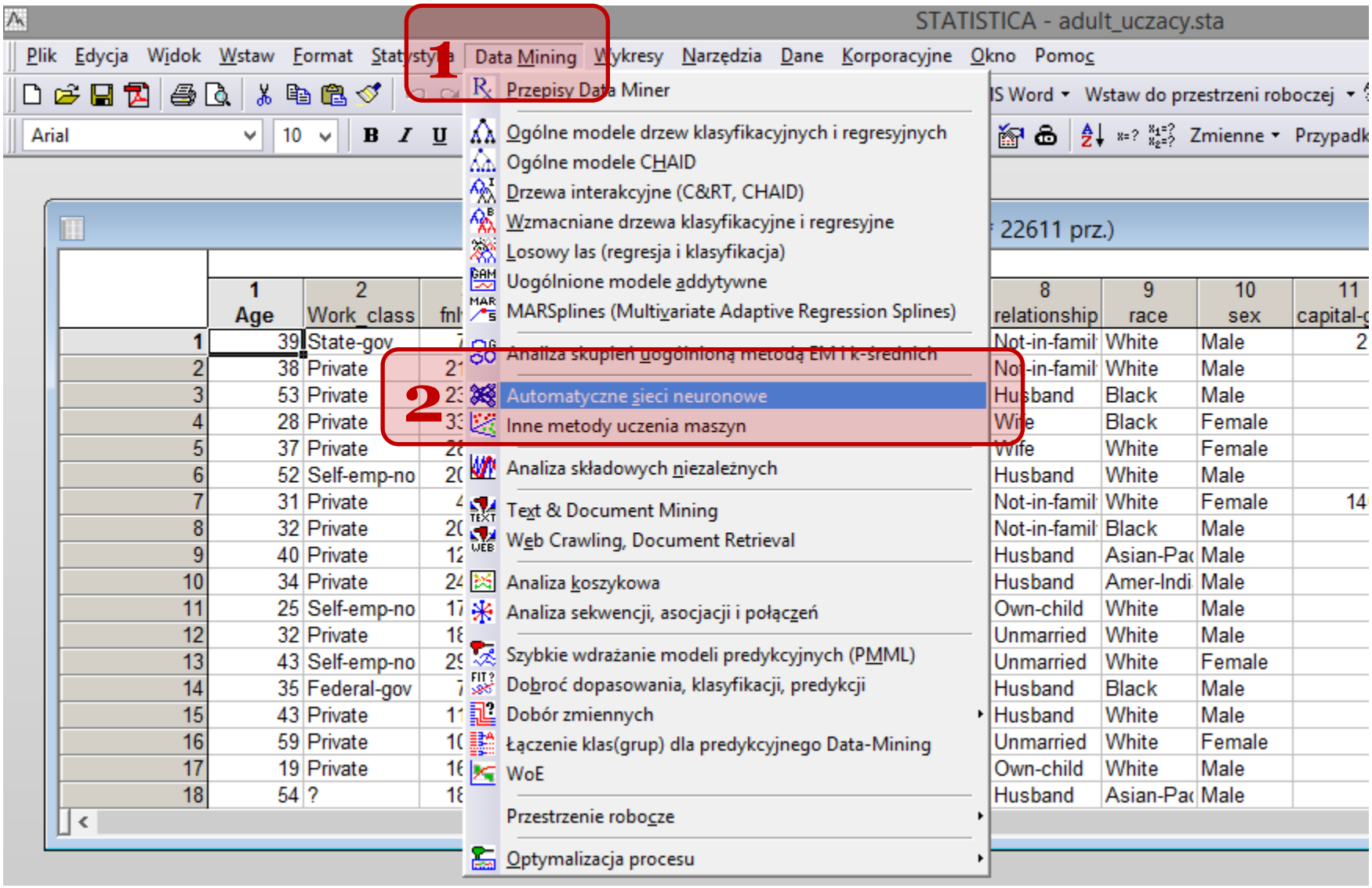
Prawidłowa procedura uczenia sieci

- Ocena jakości sieci
- Zbiór danych do uczenia sieci
- Skalowanie sygnałów
- Liczebność zbioru danych uczących
- Rozkład danych w zbiorze uczącym
- Trendy danych uczących
- Sieć nie ma zdolności do ekstrapolacji

Automatyczne Sieci Neuronowe
STATISTICA

Adult_uczacy.sta

Adult_uczacy.sta



SANN – Nowa Analiza/Wdrożenie: adult_uczacy.sta

Nowa analiza lub wdrożenie

Wdrożenie

Wdrożenie modeli z poprzednich analiz

Wczytaj pliki sieci

Nazwa pliku	ID sieci	Nazwa sieci	Aktywacja ...	Aktywacj...

Nowa analiza

Nowa analiza

3

Regresja

Klasyfikacja

Szeregi czasowe (regresja)

Szeregi czasowe (klasyfikacja)

Analiza skupień

Proszę wybrać rodzaj analizy z listy

OK

Anuluj

Opcje

Otwórz dane

SANN - Wybór danych: adult_uczacy.sta

Podstawowe | Wybór podzbiorów | Wielokrotne próbkowanie

4

Zmienne

Zmienne analizy (obecne w pliku danych)

Wyjściowe ilościowe: brak

Wyjściowe jakościowe: brak

Wejściowe ilościowe: brak

Wejściowe jakościowe: brak

Sposób tworzenia modeli

Automatyczne poszukiwanie sieci

Projekt sieci użytkownika

Wielokrotne próbkowanie (losowe, bootstrap)

OK

Anuluj

Opcje

Proszę wybrać zmienne

6

2 - Work_class	1 - Age	2 - Work_class
4 - education	3 - fnlwgt	4 - education
6 - marital-status	5 - education-num	6 - marital-status
7 - occupation	11 - capital-gain	7 - occupation
8 - relationship	12 - capital-loss	8 - relationship
9 - race	13 - hours-per-week	9 - race
10 - sex		10 - sex
14 - native-country		14 - native-country
15 - Income		15 - Income
Rozwiń / Przybliż	Rozwiń / Przybliż	Rozwiń / Przybliż
Wyjścia jakościowe:	Wejścia ilościowe:	Wejścia jakościowe:
15	1 5 11-13	2 4 6-10 14

OK

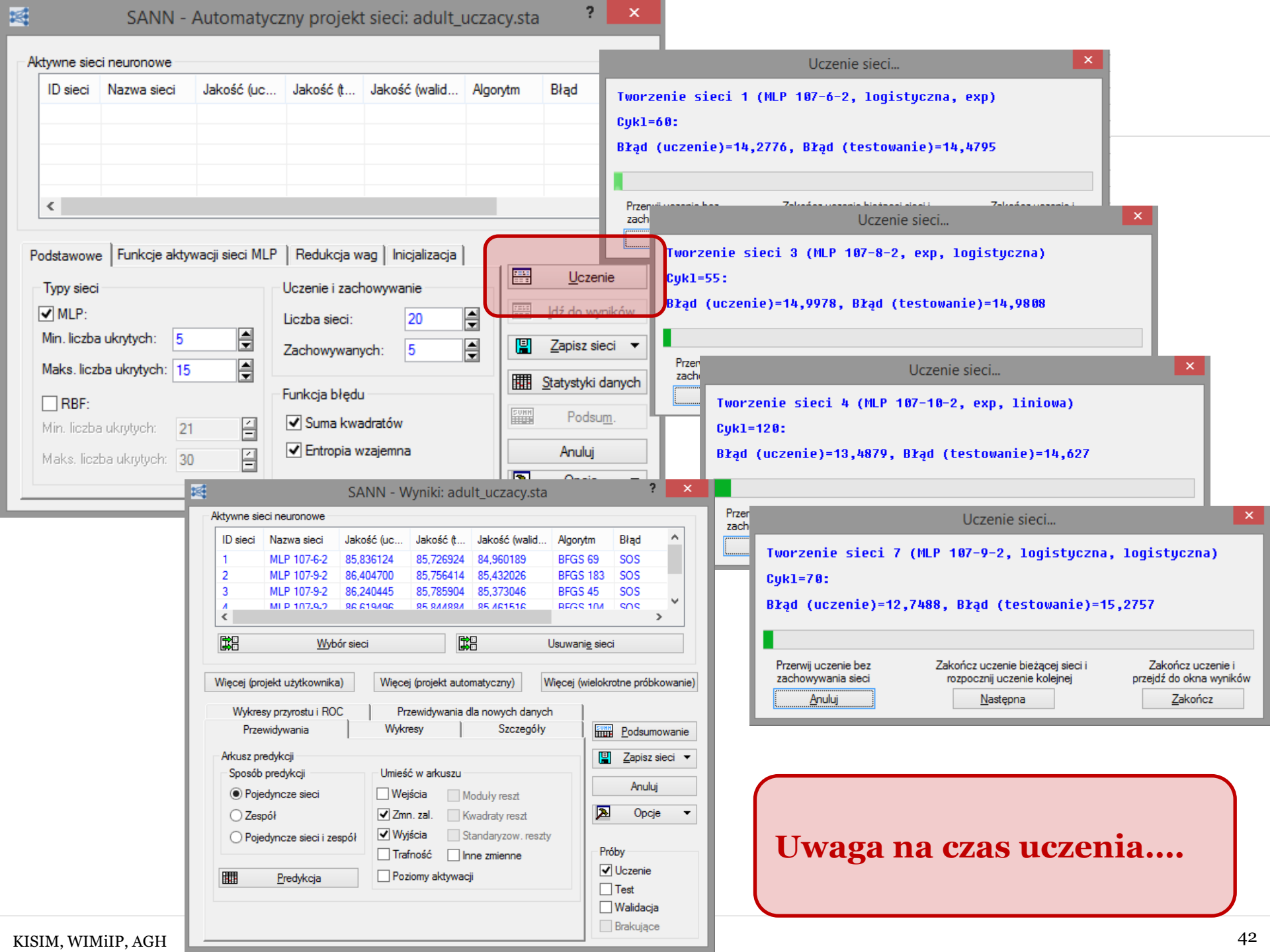
Anuluj

[Zestawy]...

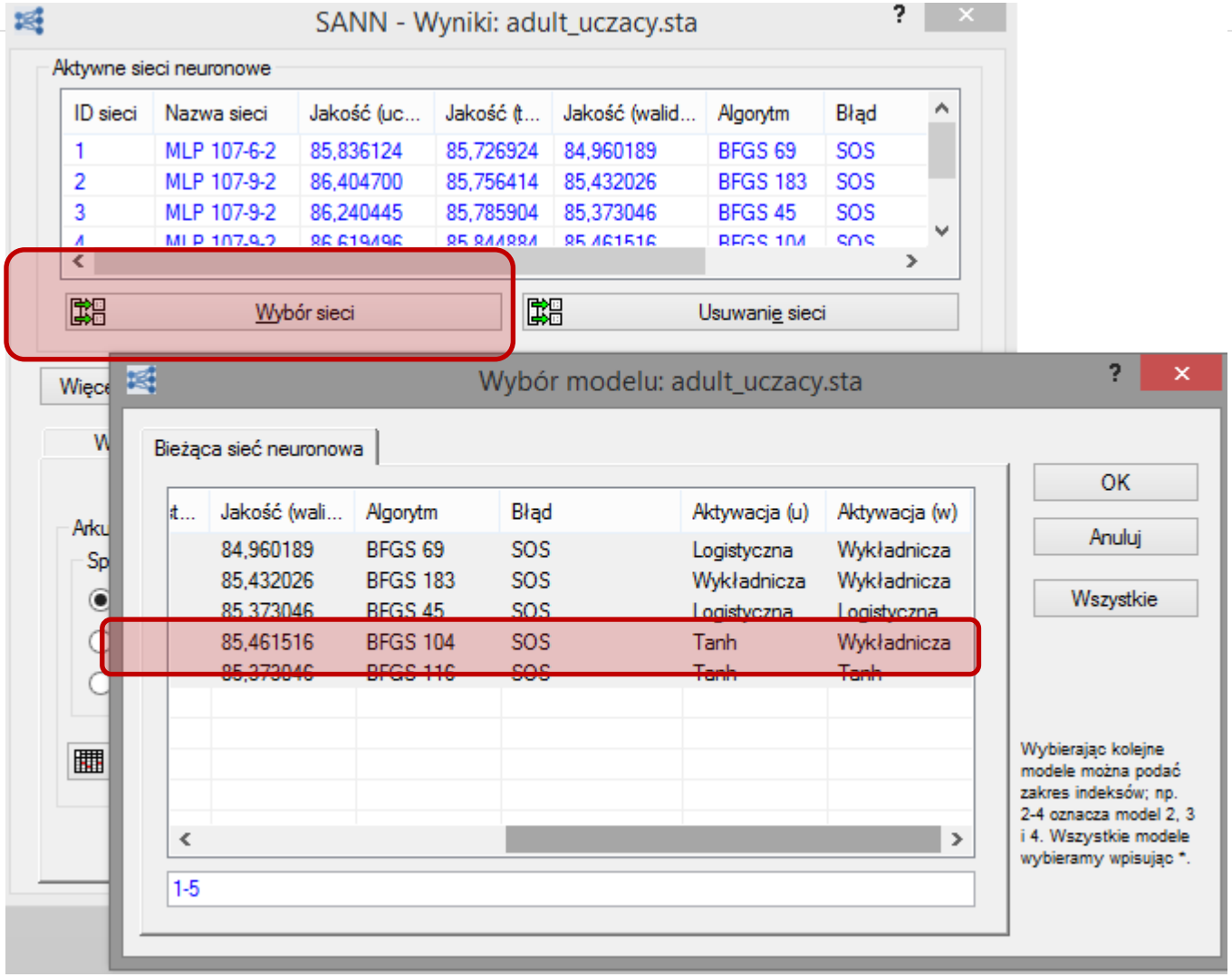
Włącz opcję "Pokazuj tylko zmienne o odpowiedniej skali" aby na listach, w zależności od potrzeby, pojawiały się tylko zmienne jakościowe albo ilościowe. Naciśnij F1 aby uzyskać więcej informacji.

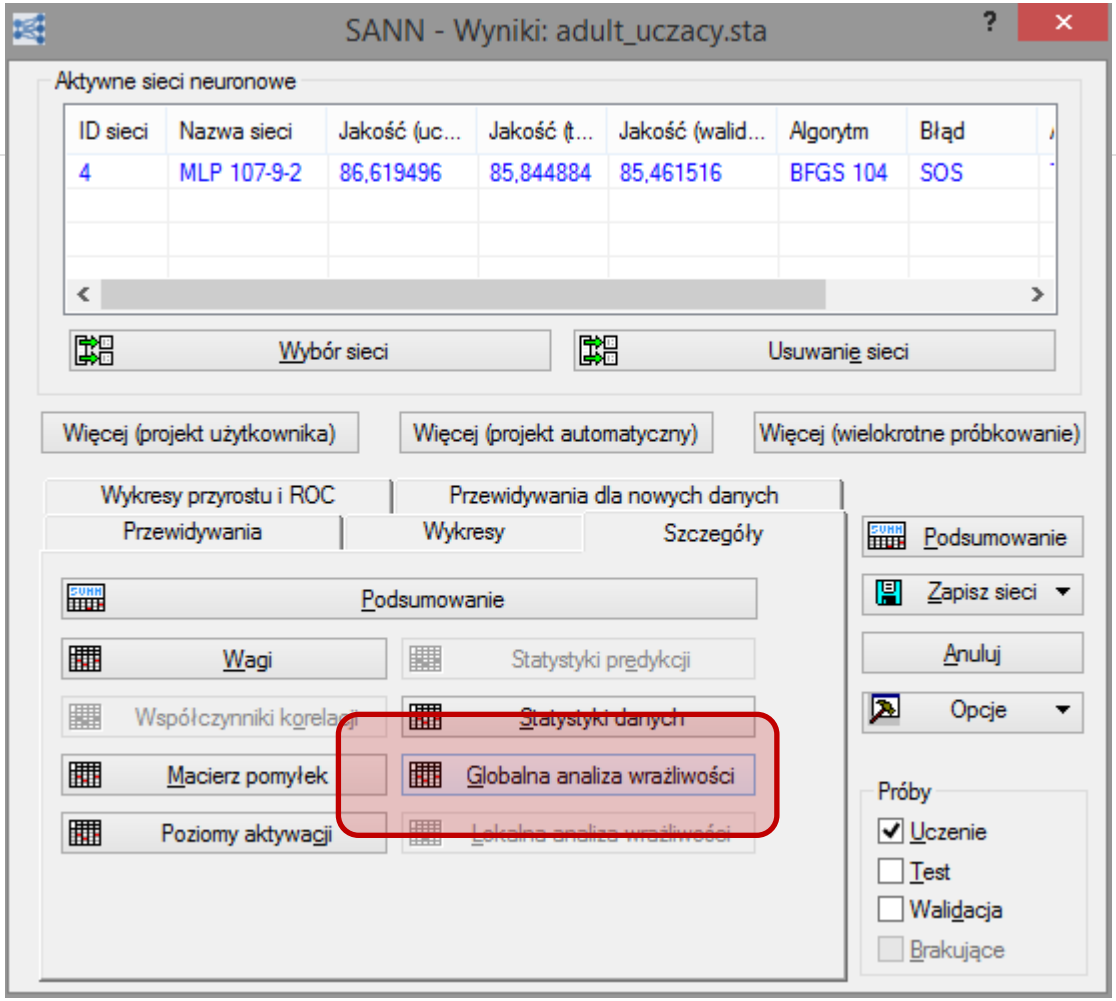
5

☑ Pokazuj tylko zmienne o odpowiedniej skali



Uwaga na czas uczenia....

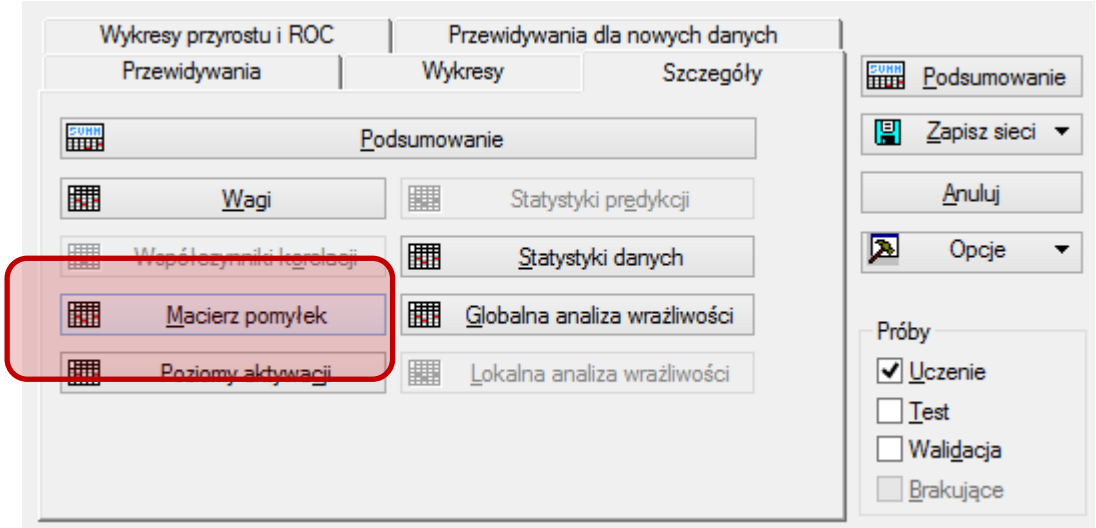

SANN - Wyniki: adult_uczacy.sta
Aktywne sieci neuronowe
ID sieci | Nazwa sieci | Jakość (uc... | Jakość (t... | Jakość (walid... | Algorytm | Błąd
1 | MLP 107-6-2 | 85,836124 | 85,726924 | 84,960189 | BFGS 69 | SOS
2 | MLP 107-9-2 | 86,404700 | 85,756414 | 85,432026 | BFGS 183 | SOS
3 | MLP 107-9-2 | 86,240445 | 85,785904 | 85,373046 | BFGS 45 | SOS
Wybór sieci
Usuwanie sieci
Wybór modelu: adult_uczacy.sta
Bieżąca sieć neuronowa
t... | Jakość (wali... | Algorytm | Błąd | Aktywacja (u) | Aktywacja (w)
84,960189 | BFGS 69 | SOS | Logistyczna | Wykładnicza
85,432026 | BFGS 183 | SOS | Wykładnicza | Wykładnicza
85,373046 | BFGS 45 | SOS | Logistyczna | Logistyczna
85,461516 | BFGS 104 | SOS | Tanh | Wykładnicza
85,373046 | BFGS 116 | SOS | Tanh | Tanh
1-5
OK
Anuluj
Wszystkie
Wybierając kolejne modele można podać zakres indeksów; np. 2-4 oznacza model 2, 3 i 4. Wszystkie modele wybieramy wpisując *.



Analiza wrażliwości (adult_uczacy.sta)
Próby: Uczenie

Sieci	relationship	marital-status	native-country	occupation	Work_class	education	capital-gain	race	Age
4.MLP 107-9-2	2,477051	1,723802	1,385571	1,260207	1,252373	1,224393	1,108506	1,076101	1,060849

education-num	hours-per-week	capital-loss	sex
1,035317	1,025169	1,016219	1,000000



Income (Podsumowanie klasyfikacji) (adult_uczacy.sta)
Próby: Uczenie

		Income-<=50K	**Income->50K**	Income-Wszystkie
4.MLP 107-9-2	Razem	12008,00	3821,000	15829,00
	Poprawne	11317,00	2394,000	13711,00
	Niepoprawne	691,00	1427,000	2118,00
	Poprawne (%)	94,25	62,654	86,62
	Niepoprawne (%)	5,75	37,346	13,38

Przewidywania | Wykresy | Szczegóły

Wykresy przyrostu i ROC | Przewidywania dla nowych danych

Podsumowanie

Zapisz sieci

Anuluj

Opcje

Liczba przewidywanych: 5

Usuń poprzednie przewidywania

Wartości na wejściu

Prześlij do arkusza

#	4.Income	Age	educati...	capital-...	capital-l...	hours-1

Wartości wejściowe

Wartości wejściowe

	Work_class	education	marital-status	occupation	relatio
Przypadek 1	Never-married	Adm-clerical	Not-in-family	White	Male
Przypadek 2	Divorced	Handlers-clea	Not-in-family	White	Male

Wykresy przyrostu i ROC | Przewidywania dla nowych danych

Przewidywania | Wykresy | Szczegóły

Podsumowanie

Zapisz sieci

Anuluj

Opcje

Arkusz predykcji

Sposób predykcji

- Pojedyncze sieci
- Zespół
- Pojedyncze sieci i zespół

Umieść w arkuszu

- [x] Wejścia
- [x] Zmn. zal.
- [x] Wyjścia
- [x] Trafność
- [ ] Poziomy aktywacji
- [ ] Moduły reszt
- [ ] Kwadraty reszt
- [ ] Standaryzow. reszty
- [ ] Inne zmienne

Próby

- [x] Uczenie
- [ ] Test

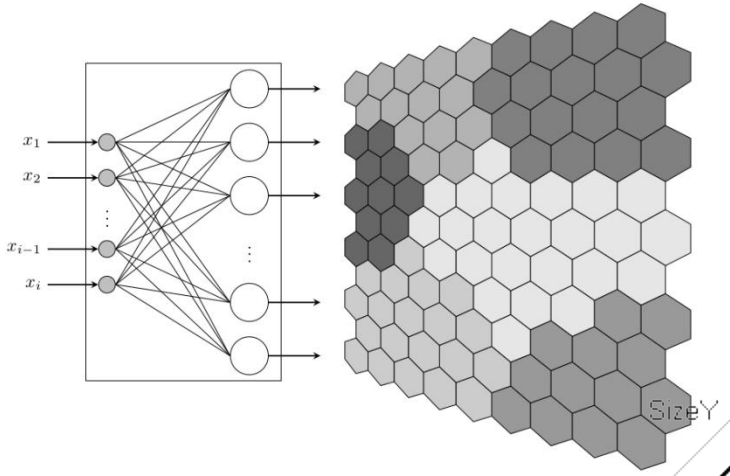
Predykcja

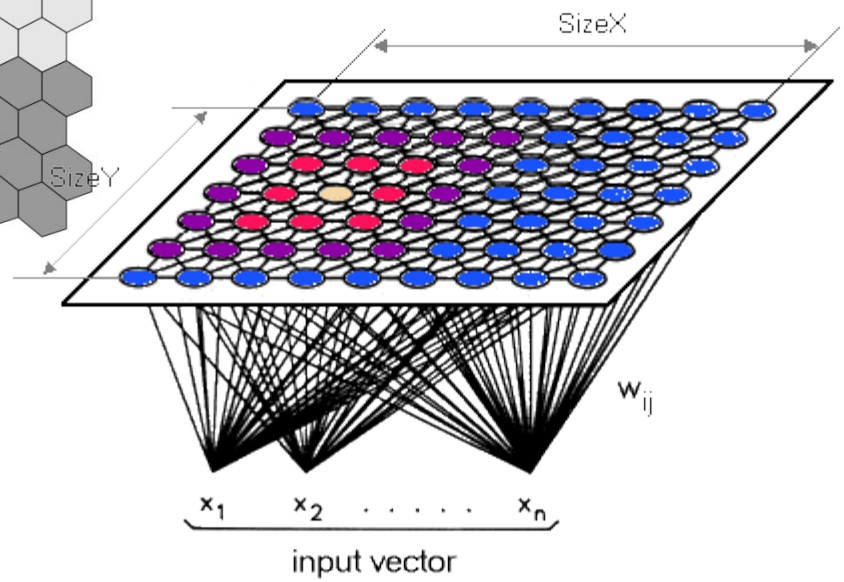
Próby: Uczenie

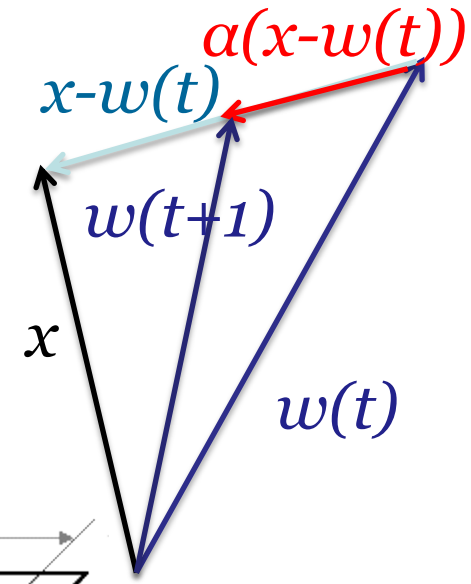
Przypadek nazwa	marital-status Wejście	occupation Wejście	relationship Wejście	race Wejście	sex Wejście	native-country Wejście	Income Zm.zal	Income - Wyjście 4. MLP 107-9-2	Income - Trafność 4. MLP 107-9-2
1	Never-married	Adm-clerical	Not-in-family	White	Male	United-States	<=50K	<=50K	Poprawne
2	Divorced	Handlers-cleaners	Not-in-family	White	Male	United-States	<=50K	<=50K	Poprawne
3	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	United-States	<=50K	<=50K	Poprawne
4	Married-civ-spouse	Prof-specialty	Wife	Black	Female	Cuba	<=50K	<=50K	Poprawne
5	Married-civ-spouse	Exec-managerial	Wife	White	Female	United-States	<=50K	>50K	Niepoprawne
6	Married-civ-spouse	Exec-managerial	Husband	White	Male	United-States	>50K	<=50K	Niepoprawne
8	Never-married	Sales	Not-in-family	Black	Male	United-States	<=50K	<=50K	Poprawne
9	Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	?	>50K	<=50K	Niepoprawne
10	Married-civ-spouse	Transport-moving	Husband	Amer-Indian-Eskimo	Male	Mexico	<=50K	<=50K	Poprawne

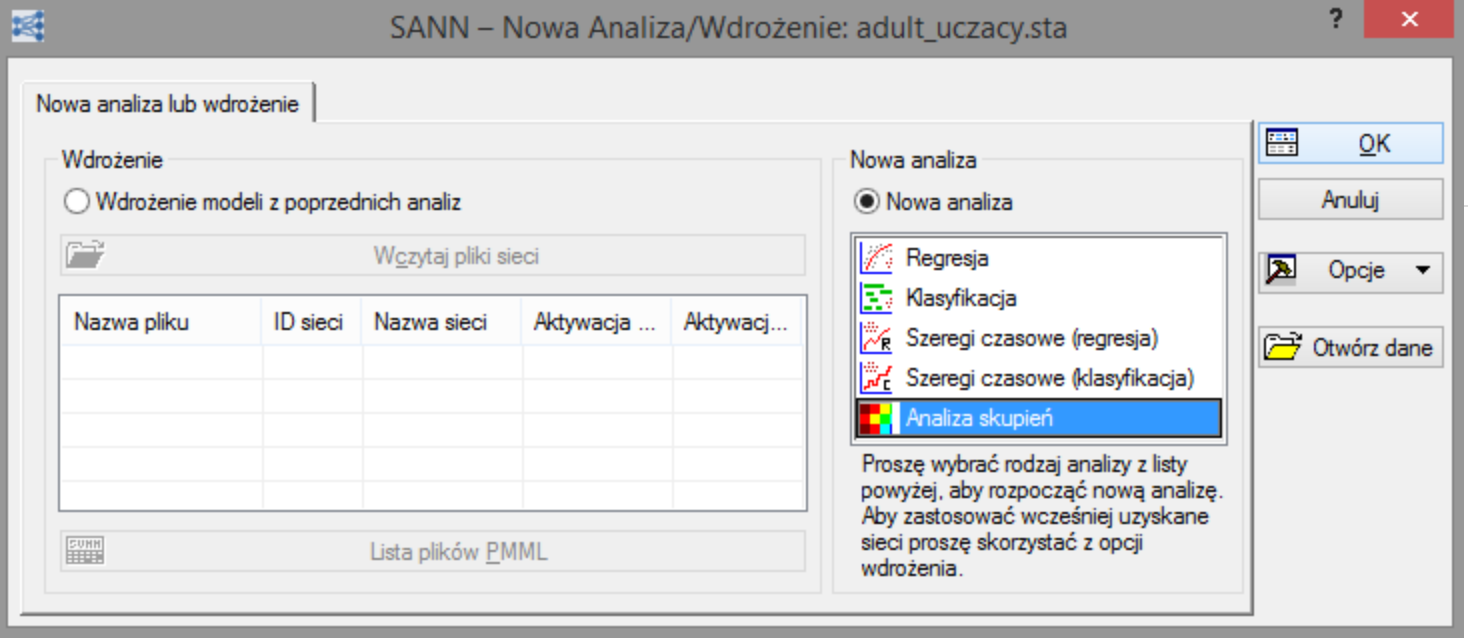
Sieci Kohonena

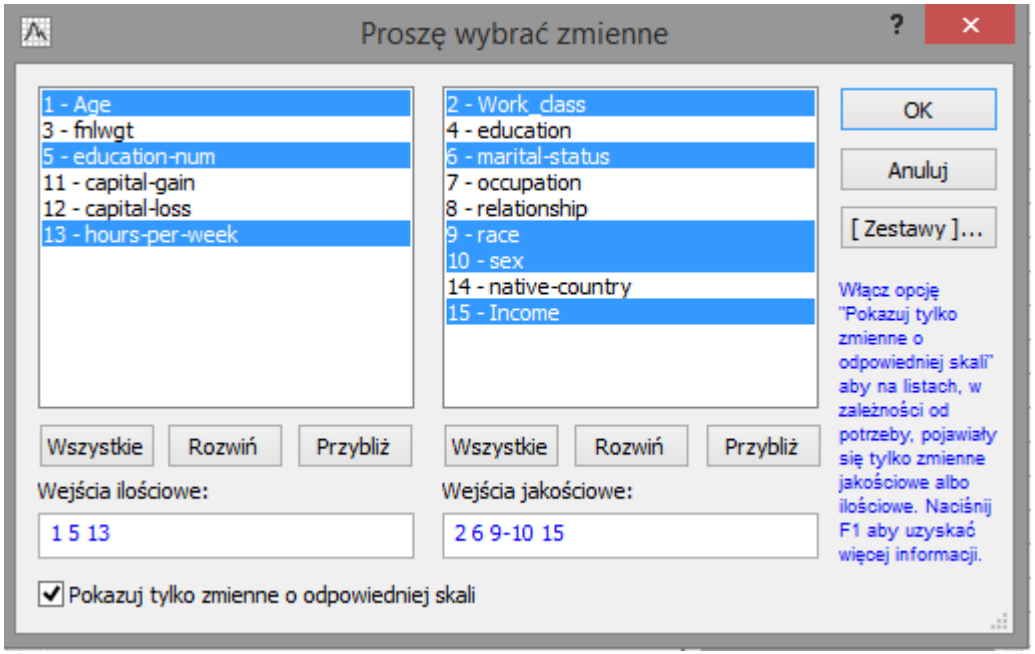
(samoorganizujące się)
(Self Organizing Maps – SOM)

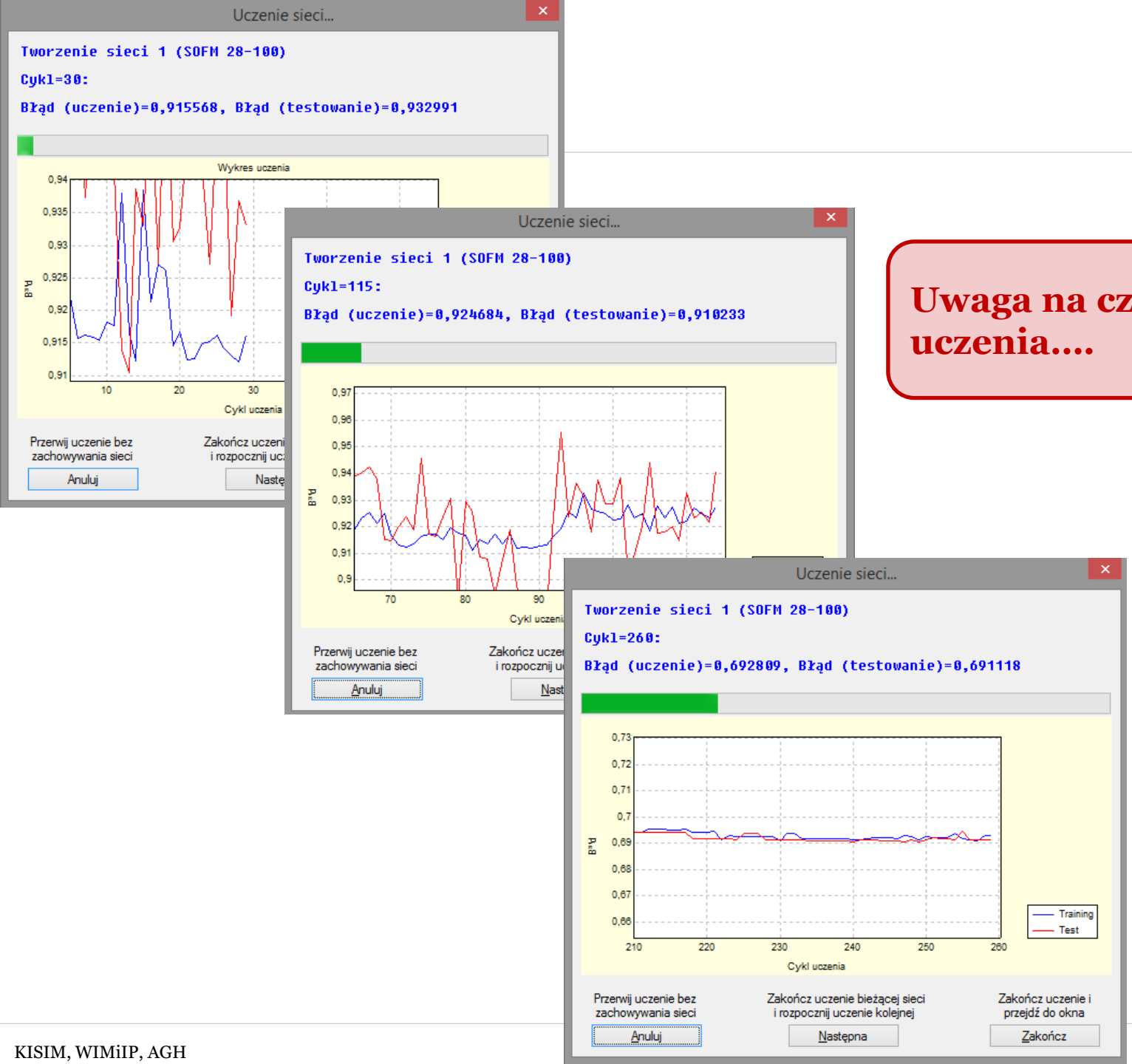







SANN – Nowa Analiza/Wdrożenie: adult_uczacy.sta
Nowa analiza lub wdrożenie
Wdrożenie
Wdrożenie modeli z poprzednich analiz
Wczytaj pliki sieci
Nazwa pliku
ID sieci
Nazwa sieci
Aktywacja ...
Aktywacj...
Lista plików PMML
Nowa analiza
Nowa analiza
Regresja
Klasyfikacja
Szeregi czasowe (regresja)
Szeregi czasowe (klasyfikacja)
Analiza skupień
Proszę wybrać rodzaj analizy z listy powyżej, aby rozpocząć nową analizę. Aby zastosować wcześniej uzyskane sieci proszę skorzystać z opcji wdrożenia.
OK
Anuluj
Opcje
Otwórz dane

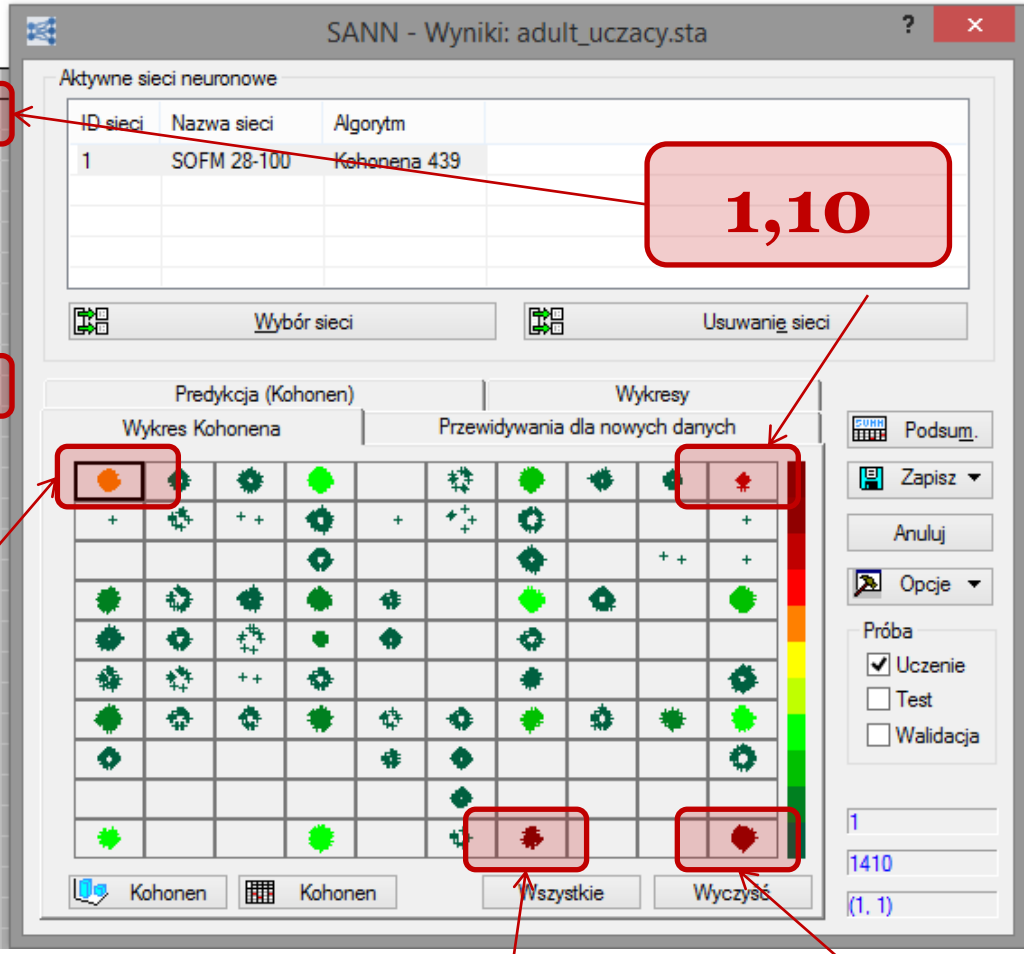

Proszę wybrać zmienne
1 - Age
3 - fnlwgt
5 - education-num
11 - capital-gain
12 - capital-loss
13 - hours-per-week
2 - Work_class
4 - education
5 - marital-status
7 - occupation
8 - relationship
9 - race
10 - sex
14 - native-country
15 - Income
OK
Anuluj
[Zestawy]...
Włącz opcję "Pokazuj tylko zmienne o odpowiedniej skali" aby na listach, w zależności od potrzeby, pojawiały się tylko zmienne jakościowe albo ilościowe. Naciśnij F1 aby uzyskać więcej informacji.
Wszystkie
Rozwiń
Przybliż
Wszystkie
Rozwiń
Przybliż
Wejścia ilościowe:
1 5 13
Wejścia jakościowe:
2 6 9-10 15
Pokazuj tylko zmienne o odpowiedniej skali



Uwaga na czas uczenia....

Arkusz liczności (adult_uczacy.sta)
Sieć: 1.SOFM 28-100
Próby: Uczenie

	1	2	3	4	5	6	7	8	9	10
1	1410	34	66	695	0	12	535	42	84	1741
2	1	10	2	47	1	5	34	0	0	1
3	0	0	0	37	0	0	173	0	2	1
4	330	24	61	350	19	0	661	48	0	530
5	146	28	10	214	47	0	21	0	0	0
6	25	10	2	19	0	0	63	0	0	99
7	336	16	16	400	8	23	442	24	262	711
8	47	0	0	0	25	48	0	0	0	47
9	0	0	0	0	0	67	0	0	0	0
10	798	0	0	689	0	7	2137	0	0	2086



1,1

1,10

10,7

10,10

Arkusz Kohonena dla wybranych komórek (adult_uczacy.sta)
Liczba wybranych przypadków: 2137, Liczba wybranych komórek: 1
Próby: Uczenie

10,7

Przypadek nazwa	Położenie neuronu	ID neuronu	Aktywacja	Age	education-num	hours-per-week	Work_class	marital-status	race	sex	Income
6139	(10, 7)	97	0,107647	47	9						
6147	(10, 7)	97	0,463776	32	9						
6161	(10, 7)	97	0,110170	36	9						
6184	(10, 7)	97	0,177701	30	9						
6197	(10, 7)	97	0,254025	24	9						
6223	(10, 7)	97	0,333574	19	9						
6224	(10, 7)	97	0,098799	46	9						
6247	(10, 7)	97	0,444098	60	3						
6252	(10, 7)	97	0,392284	53	3						
6255	(10, 7)	97	0,203770	31	9						
6259	(10, 7)	97	0,174438	53	9						
6278	(10, 7)	97	0,118129	39	10						
6291	(10, 7)	97	0,456679	23	3						
6302	(10, 7)	97	0,372806	27	13	40,00	Private	Married-civ-spouse	White	Male	<=50K
6314	(10, 7)	97	0,364129	54	11	20,00	Private	Married-civ-spouse	White	Male	<=50K
6315	(10, 7)	97	0,162050	36	9	60,00	Private	Married-civ-spouse	White	Male	<=50K
6324	(10, 7)	97	0,160349	33	10	45,00	Private	Married-civ-spouse	White	Male	<=50K
6332	(10, 7)	97	0,326479	21	6						
6339	(10, 7)	97	0,101049	37	9						
6358	(10, 7)	97	0,405340	66	7						
6363	(10, 7)	97	0,388225	25	13						
6376	(10, 7)	97	0,662576	72	13						
6381	(10, 7)	97	0,141536	37	10						
6421	(10, 7)	97	0,138422	36	7						
6428	(10, 7)	97	0,174438	53	9						

Wybierz tylko nazwy zmiennych
Statystyki bloku danych
Wykresy bloku danych
Wykresy danych wejściowych
Wytnij Ctrl+X
Kopiuj Ctrl+C
Kopiuj z nagłówkami
Wklej Ctrl+V
Wklej specjalnie...
Dodaj zmienne...
Usuń zmienne...

Dla kolumn
Średnie
Mediany
Odch.std.
N ważnych
Sumy
Minima
Maksima
25%
75%
Wszystko

Przypadek nazwa	1 Age	2 education-num	3 hours-per-week
ŚREDNIA przyp. 1-2137	40,529	9,10481984	42,6518484
MEDIANA przyp. 1-2137	38	9	40
ODCH.STD. przyp. 1-2137	12,17	2,50902121	9,61727766
N_WAŻNYCH przyp. 1-2137	2137	2137	2137
SUMA przyp. 1-2137	86611	19457	91147
MINIMUM przyp. 1-2137	18	1	2
MAKSIMUM przyp. 1-2137	90	16	99
_25% przyp. 1-2137	31	9	40
_75% przyp. 1-2137	49	10	45

Dane: Arkusz27 (5 zmn. * 2137 prz.)

	Work_class	marital-status	race	sex	Income
10	Private	Married-civ-spouse	-Eskimo	Male	<=50K
15	Private	Married-civ-spouse	White	Male	<=50K
25	Private	Married-civ-spouse	White	Male	<=50K
29	Private	Married-civ-spouse	White	Male	<=50K
39	Private	Married-civ-spouse	White	Male	<=50K
40	Private	Married-civ-spouse	White	Male	<=50K
41	Private	Married-civ-spouse	White	Male	<=50K
44	Private	Married-civ-spouse	White	Male	<=50K
50	Private	Married-civ-spouse	White	Male	<=50K
52	Private	Married-civ-spouse	White	Male	<=50K
72	Private	Married-civ-spouse	White	Male	<=50K
77	Private	Married-civ-spouse	White	Male	<=50K
87	Private	Married-civ-spouse	White	Male	<=50K
122	Private	Married-civ-spouse	White	Male	<=50K
131	Private	Married-civ-spouse	White	Male	<=50K
144	Private	Married-civ-spouse	White	Male	<=50K

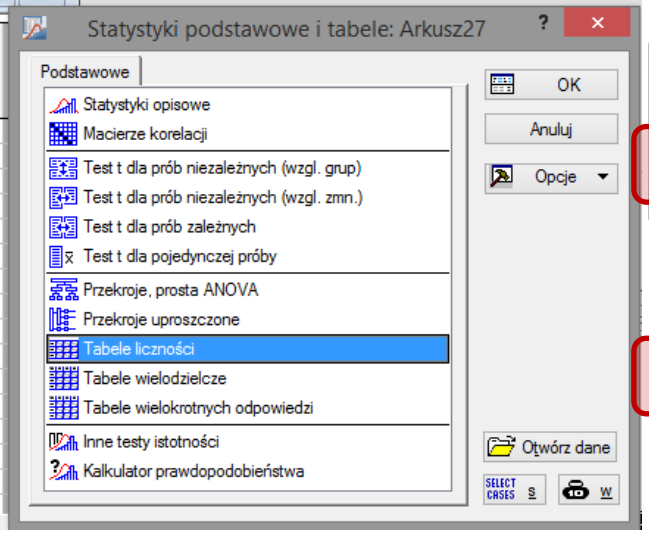


Tabela liczności: Work_class (Arkus:

Klasa	Liczba	Skumulow. Liczba	Procent
Private	2137	2137	100,000
Braki	0	2137	0,000

Tabela liczności: sex (Arkusz27)

Klasa	Liczba	Skumulow. Liczba	Procent
Male	2137	2137	100,0000
Braki	0	2137	0,0000

Tabela liczności: marital

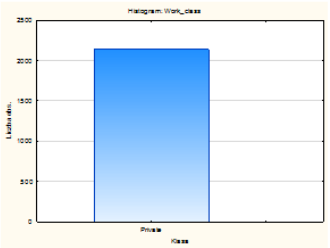
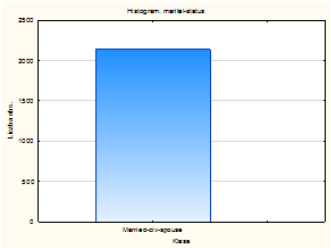
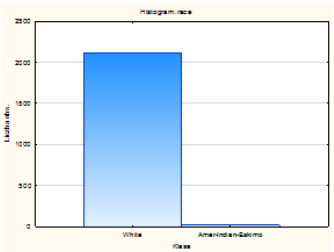
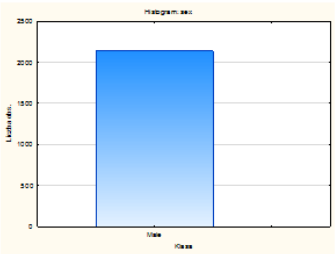
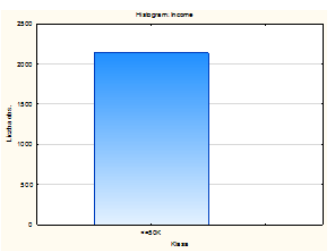
Klasa	Liczba	Skumulow. Liczba
Married-civ-spouse	2137	213
Braki	0	213

Tabela liczności: race (

Klasa	Liczba	Skumulow Liczba
White	2118	21
Amer-Indian-Eskimo	19	21
Braki	0	21

Tabela liczności: Income (Arkusz27)

Klasa	Liczba	Skumulow. Liczba	Procent
<=50K	2137	2137	100,0000
Braki	0	2137	0,0000

Arkusz Kohonena dla wybranych komórek (adult_uczacy.sta)
Liczba wybranych przypadków: 2086, Liczba wybranych komórek: 1
Próby: Uczenie

10,10

Przypadek nazwa	Położenie neuronu	ID neuronu	Aktywacja	Age	education-num	hours-per-week	Work_class	marital-status	race	sex	Income
9	(10, 10)	100	1,2	40	11	40	Private	Married-civ-spouse	:-Islander	Male	>50K
23	(10, 10)	100	0,3	31	10	38	Private	Married-civ-spouse	White	Male	>50K
38	(10, 10)	100	0,3	43	10	40	Private	Married-civ-spouse	White	Male	>50K
59	(10, 10)	100	0,3	37	10	48	Private	Married-civ-spouse	White	Male	>50K
62	(10, 10)	100	0,6	76	14						
92	(10, 10)	100	1,2	49	10						
124	(10, 10)	100	0,6	40	4						
155	(10, 10)	100	1,4	28	15	55	Private	Never-married	White	Male	>50K
160	(10, 10)	100	0,3	40	13	40	Private	Married-civ-spouse	White	Male	>50K
165	(10, 10)	100	0,4	58	9	45	Private	Married-civ-spouse	White	Male	>50K
184	(10, 10)	100	0,3	44	10	40	Private	Married-civ-spouse	White	Male	>50K
194	(10, 10)	100	0,4	41	15	60	Private	Married-civ-spouse	White	Male	>50K
205	(10, 10)	100	0,3	31	9	45	Private	Married-civ-spouse	White	Male	>50K

education-num

Przypadek nazwa	1 Age	2 education-num	3 hours-per-week
ŚREDNIA przyp. 1-2086	43,118	11,4237776	46,6514861

Tabela liczności: Work_class (Arkus:

Klasa	Liczba	Skumulow. Liczba	Procent
Private	2137	2137	100,000
Braki	0	2137	0,000

Tabela liczności: sex (Arkusz27)

Klasa	Liczba	Skumulow. Liczba	Procent
Male	2137	2137	100,0000
Braki	0	2137	0,0000

Tabela liczności: Income (Arkusz27)

Klasa	Liczba	Skumulow. Liczba	Procent
<=50K	51	51	2,38652
>50K	2086	2137	97,61348
Braki	0	2137	0,00000

Tabela liczności: ma

Klasa	Liczba	Skumul Liczb
Never-married	113	
Divorced	75	
Married-civ-spouse	1910	
Married-AF-spouse	1	
Separated	22	
Married-spouse-absent	6	
Widowed	10	
Braki	0	

Tabela liczności: race

Klasa	Liczba	Skumulow Liczba
White	1993	19
Black	65	20
Asian-Pac-Islander	66	21
Amer-Indian-Eskimo		
Other		
Braki		

race	marital-status	Income <=50K	Income >50K	Wiersz Razem
White	Never-married	0	108	108
White	Divorced	0	72	72
White	Married-civ-spouse	51	1723	1774
White	Married-AF-spouse	0	1	1
White	Separated	0	22	22
White	Married-spouse-absent	0	6	6
White	Widowed	0	10	10
Ogół		51	1942	1993

1,1

Przypadek nazwa	1 Age	2 education-num	3 hours-per-week
ŚREDNIA przyp. 1-1410	26,9595745	10,1035461	34,4404255
MEDIANA przyp. 1-1410	24	10	40
ODCH.STD. przyp. 1-1410	9,45966356	2,29401701	11,8861256
N_WAŻNYCH przyp. 1-1410	1410	1410	1410
SUMA przyp. 1-1410	38013	14246	48561
MINIMUM przyp. 1-1410	17	1	2
MAKSIMUM przyp. 1-1410	90	16	80
_25% przyp. 1-1410	20	9	25
_75% przyp. 1-1410	30	12	40

	Tabela liczności: Work_class (Arkus:		
Klasa	Liczba	Skumulow. Liczba	Procent
Private	1461	1461	68,3668
Braki	676	2137	31,6331

	Tabela liczności: marital	
Klasa	Liczba	Skumulow. Liczba
Never-married	1410	141
Married-civ-spouse	51	146
Braki	676	213

	Tabela liczności: race (Arkusz27)		
Klasa	Liczba	Skumulow. Liczba	Procent
White	1461	1461	68,36687
Braki	676	2137	31,63313

	Tabela liczności: sex (Arkusz27)		
Klasa	Liczba	Skumulow. Liczba	Procent
Male	51	51	2,3865
Female	1410	1461	65,9803
Braki	676	2137	31,6331

	Tabela liczności: Income (Arkusz27)		
Klasa	Liczba	Skumulow. Liczba	Procent
<=50K	1418	1418	66,35470
>50K	43	1461	2,01217
Braki	676	2137	31,63313

1,10

Przypadek nazwa	1 Age	2 education-num	3 hours-per-week
ŚREDNIA przyp. 1-1741	26,7512924	9,60310167	37,8230902
MEDIANA przyp. 1-1741	24	9	40
ODCH.STD. przyp. 1-1741	8,75796221	2,34112941	11,9698739
N_WAŻNYCH przyp. 1-1741	1741	1741	1741
SUMA przyp. 1-1741	46574	16719	65850
MINIMUM przyp. 1-1741	17	1	1
MAKSIMUM przyp. 1-1741	90	16	99
_25% przyp. 1-1741	21	9	32
_75% przyp. 1-1741	30	10	40

	Tabela liczności: Work_class (Arkus:		
Klasa	Liczba	Skumulow. Liczba	Procent
Private	1741	1741	81,4693
Braki	396	2137	18,5306

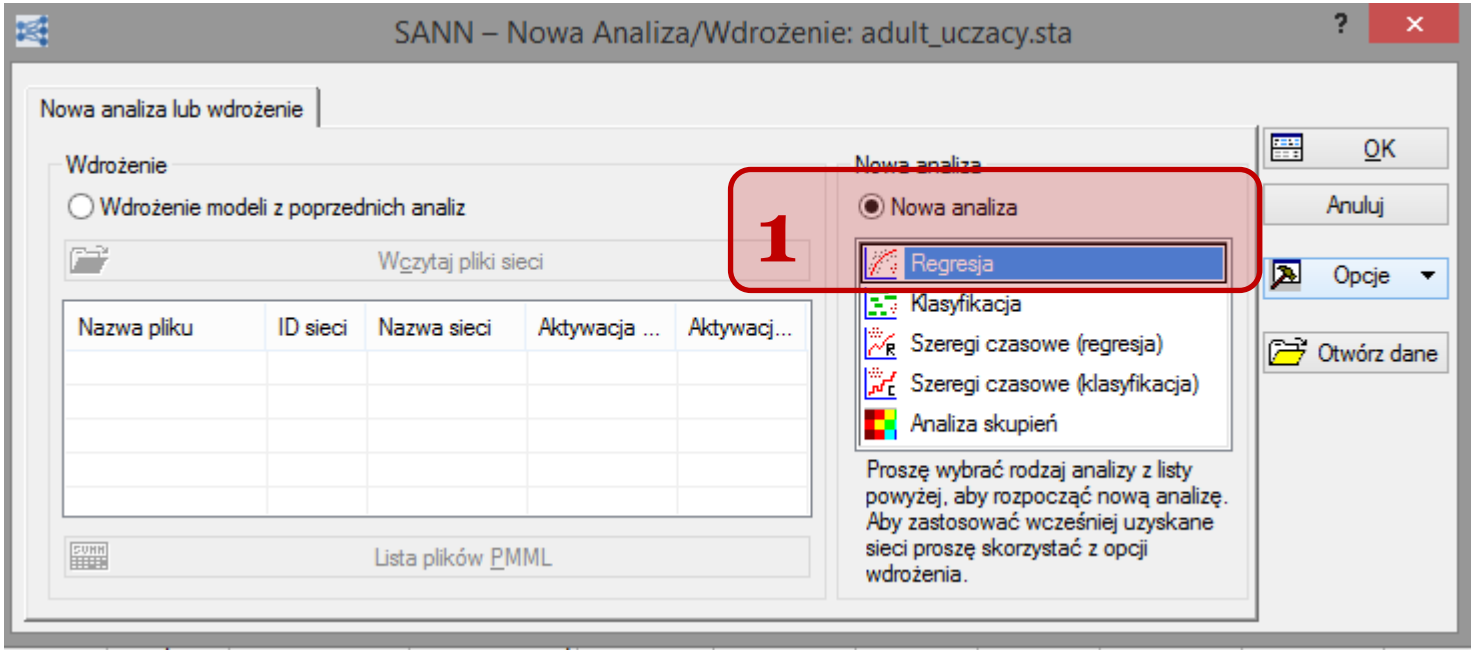
	Tabela liczności: marital-stat		
Klasa	Liczba	Skumulow. Liczba	P
Never-married	1741	1741	
Braki	396	2137	

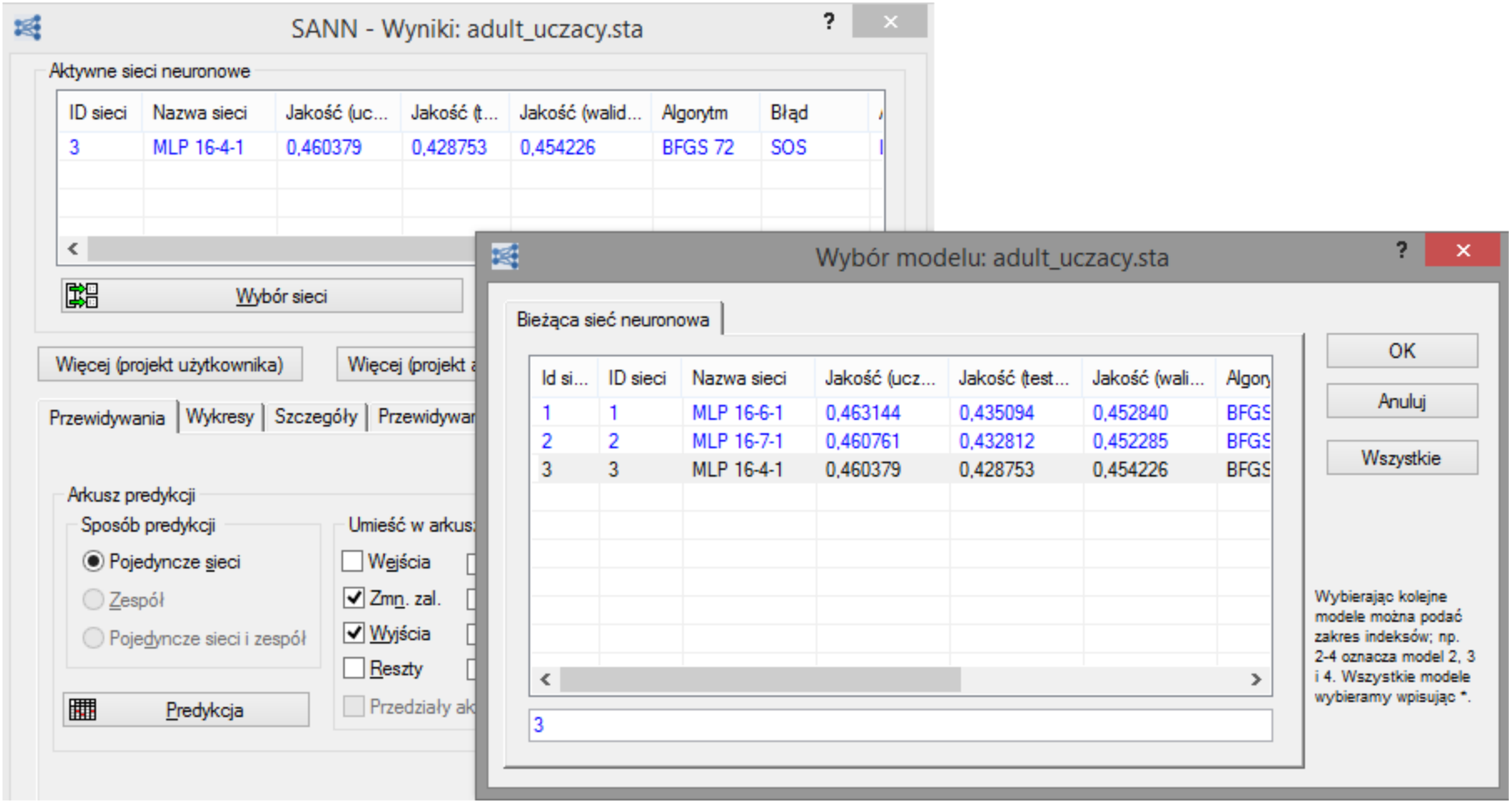
	Tabela liczności: race (Arkusz27)		
Klasa	Liczba	Skumulow. Liczba	Procent
White	1741	1741	81,46935
Braki	396	2137	18,53065

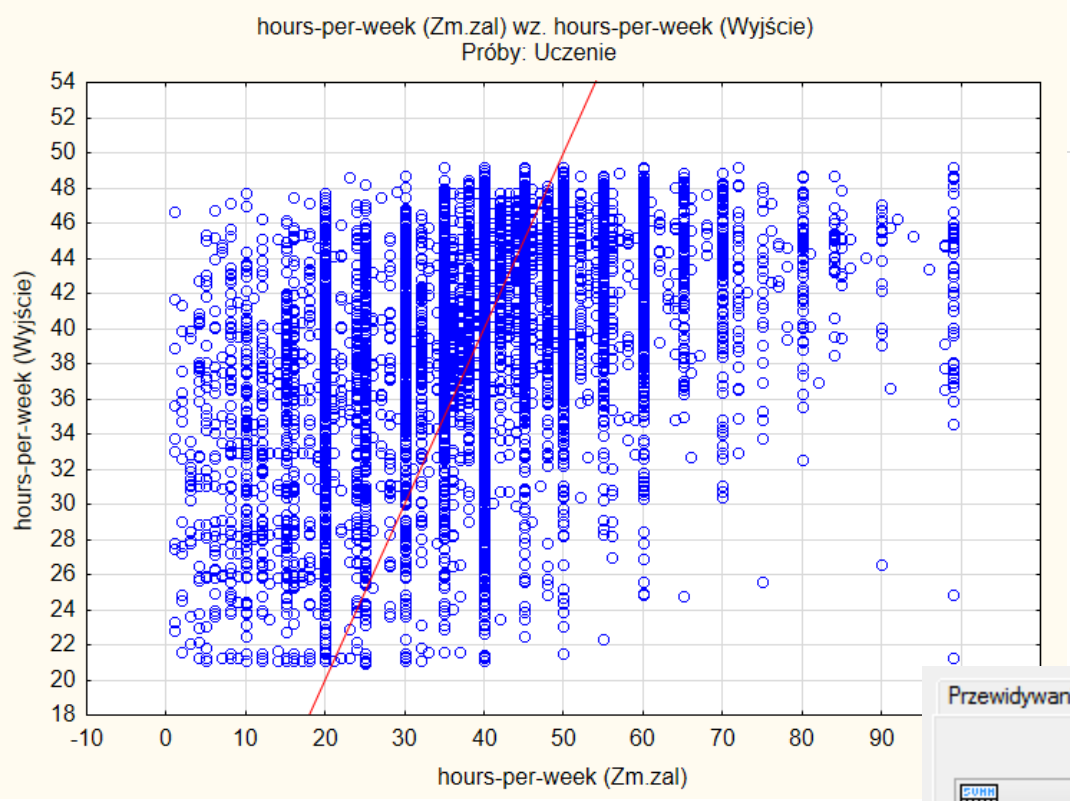
	Tabela liczności: sex (Arkusz27)		
Klasa	Liczba	Skumulow. Liczba	Procent
Male	1741	1741	81,46935
Braki	396	2137	18,53065

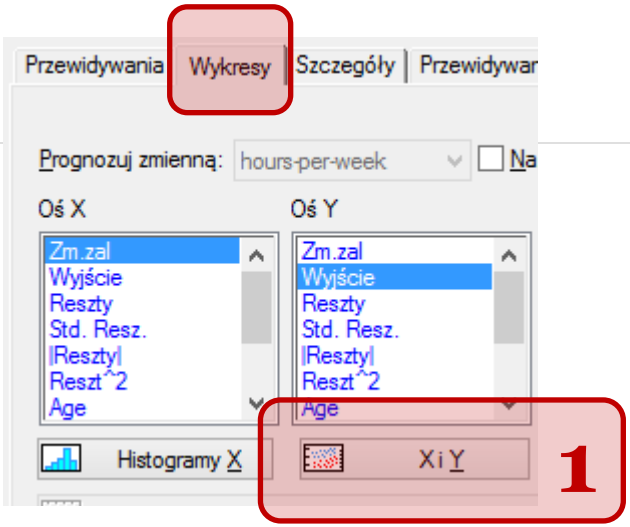
	Tabela liczności: Income (Arkusz27)		
Klasa	Liczba	Skumulow. Liczba	Procent
<=50K	1739	1739	81,37576
>50K	2	1741	0,09359
Braki	396	2137	18,53065

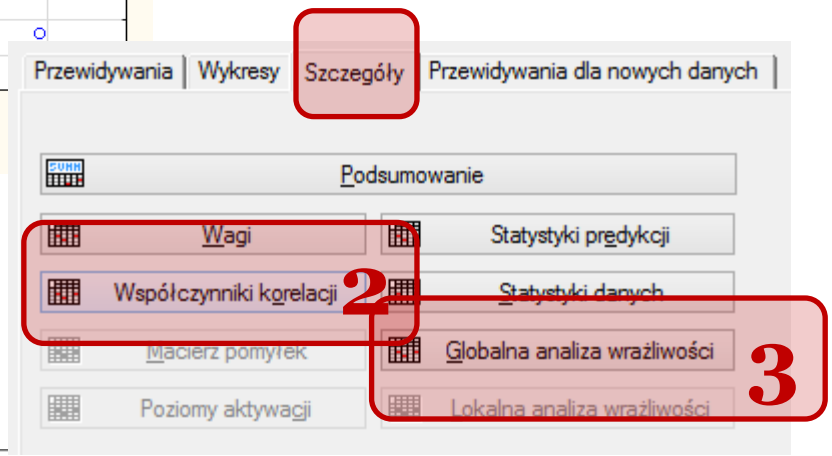
Regresja


SANN – Nowa Analiza/Wdrożenie: adult_uczacy.sta
Nowa analiza lub wdrożenie
Wdrożenie
Wdrożenie modeli z poprzednich analiz
Wczytaj pliki sieci
Nazwa pliku
ID sieci
Nazwa sieci
Aktywacja ...
Aktywacj...
Lista plików PMML
Nowa analiza
Nowa analiza
1
Regresja
Klasyfikacja
Szeregi czasowe (regresja)
Szeregi czasowe (klasyfikacja)
Analiza skupień
Proszę wybrać rodzaj analizy z listy powyżej, aby rozpocząć nową analizę. Aby zastosować wcześniej uzyskane sieci proszę skorzystać z opcji wdrożenia.
OK
Anuluj
Opcje
Otwórz dane


SANN - Wyniki: adult_uczacy.sta
Aktywne sieci neuronowe
ID sieci
Nazwa sieci
Jakość (uc...
Jakość (t...
Jakość (walid...
Algorytm
Błąd
3
MLP 16-4-1
0,460379
0,428753
0,454226
BFGS 72
SOS
Wybór sieci
Więcej (projekt użytkownika)
Więcej (projekt a
Przewidywania
Wykresy
Szczegóły
Przewidywar
Arkusz predykcji
Sposób predykcji
Pojedyncze sieci
Zespół
Pojedyncze sieci i zespół
Umieść w arkusz
Wejścia
Zmn. zal.
Wyjścia
Reszty
Przedziały ak
Predykcja
Wybór modelu: adult_uczacy.sta
Bieżąca sieć neuronowa
Id si...
ID sieci
Nazwa sieci
Jakość (ucz...
Jakość (test...
Jakość (wali...
Algory
1
1
MLP 16-6-1
0,463144
0,435094
0,452840
BFGS
2
2
MLP 16-7-1
0,460761
0,432812
0,452285
BFGS
3
3
MLP 16-4-1
0,460379
0,428753
0,454226
BFGS
3
OK
Anuluj
Wszystkie
Wybierając kolejne modele można podać zakres indeksów; np. 2-4 oznacza model 2, 3 i 4. Wszystkie modele wybieramy wpisując *.







	Współczynniki korelacji (adult_uczacy.sta)		
	hours-per-week Uczenie	hours-per-week Test	hours-per-week Walidacja
3.MLP 16-4-1	0,460379	0,428753	0,454226

	Analiza wrażliwości (adult_uczacy.sta) Próby: Uczenie				
Sieci	**Age**	marital-status	race	education-num	sex
3.MLP 16-4-1	1,223487	1,121491	1,027899	1,008001	1,000000

1 hours-per-week	2 hours-per-week - Wyjście
40,00000	44,31106
40,00000	44,57164
40,00000	41,29789
40,00000	36,97640
40,00000	38,15631
45,00000	45,12844
50,00000	42,86623
40,00000	44,15577
45,00000	43,26563
35,00000	40,14851
40,00000	43,34173
40,00000	43,91360
40,00000	38,79966
40,00000	30,74273
40,00000	44,94326
40,00000	34,84304
25,00000	37,99742
38,00000	45,24695
40,00000	47,19996
50,00000	44,85334
35,00000	40,14851
38,00	
40,00	
43,00	
60,00	
55,00	
60,00	

Dobroć dopasowania w regresji i klasyfikacji: Ar...

Zmienne: Typ zmiennej: Ilościowa / Jakościowa (klasyfikacja)

OK

Anuluj

Obserwowane: brak

Przewidywane: brak

Podstawowe | Więcej

Statystyki ilośc. zmn. zal.
- Średni kwadrat reszt
- Średnie odchylenie
- Względny błąd kwadr.
- Względne odchylenie
- Współczynnik korelacji

Statystyki jak. zmn. zal.
- Chi kwadrat
- G kwadrat
- Procent niezgodności (błędnych klasyfikacji)

2

Proszę wybrać zmienne dla testu dobroci dopasowania

1 - hours-per-week
2 - hours-per-week - Wyjście

1 - hours-per-week
2 - hours-per-week - Wyjście

OK

Anuluj

[Zestawy]...

Włącz opcję "Pokazuj tylko zmienne o odpowiedniej skali" aby na listach, w zależności od potrzeby, pojawiały się tylko zmienne jakościowe albo ilościowe. Naciśnij F1 aby uzyskać więcej informacji.

Wszystkie | Rozwiń | Przybliż

Zmienna z wartościami obserwowanymi: 1

Wszystkie | Rozwiń | Przybliż

Zmienne z wartościami: 2

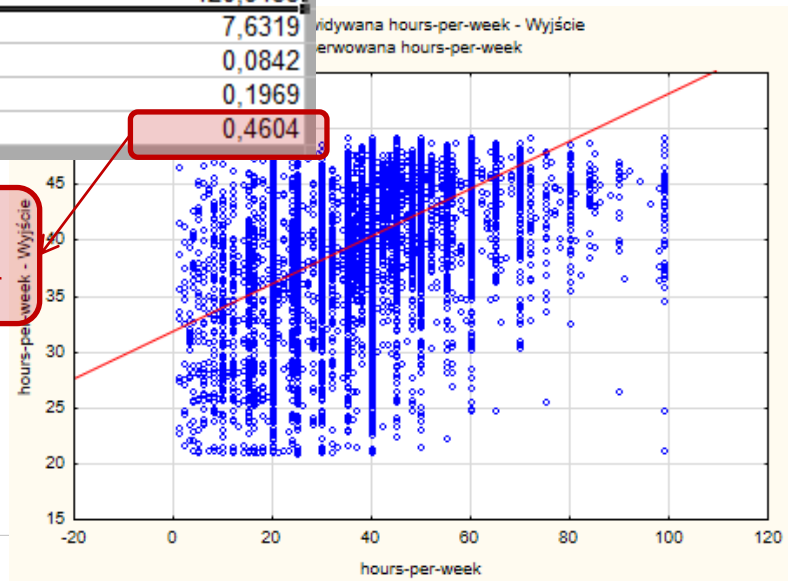
3

Data Mining | Wykresy | Narzędzia | Dane | Korporacyjne | Ol

- Przepisy Data Miner
- Ogólne modele drzew klasyfikacyjnych i regresyjnych
- Ogólne modele CHAID
- Drzewa interakcyjne (C&RT, CHAID)
- Wzmacniane drzewa klasyfikacyjne i regresyjne
- Losowy las (regresja i klasyfikacja)
- Uogólnione modele addytywne
- MARSplines (Multivariate Adaptive Regression Splines)
- Analiza skupień uogólnioną metodą EM i k-średnich
- Automatyczne sieci neuronowe
- Inne metody uczenia maszyn
- Analiza składowych niezależnych
- Text & Document Mining
- Web Crawling, Document Retrieval
- Analiza koszykowa
- Analiza sekwencji, asocjacji i połączeń
- Szybkie wdrażanie modeli predykcyjnych (PMML)
- Dobroć dopasowania, klasyfikacji, predykcji
- Dobór zmiennych
- Łączenie klas(grup) dla predykcyjnego Data Mining
- WoE
- Przestrzenie robocze
- Optymalizacja procesu

1

Podsumowanie dobroci dopasow
Zmienna obserwowana: hours-p

	hours-per-week - Wyjście
Średnia kwadratów reszt	120,9135
Średni błąd bezwzględny	7,6319
Względny średni błąd kwadratowy	0,0842
Względne odchylenie przeciętne	0,1969
Współczynnik korelacji	0,4604

$R^2=0,21$