

CW B3 Sieci neuronowe. Wykorzystanie sztucznych sieci neuronowych (MLP, RBF oraz Kohonena) w zadaniu klasyfikacji oraz porównanie ich skuteczności (rozpoznawanie płci osoby ankietowanej testem psychologicznym). Program Statistica

Ankieta na podstawie której powstała baza danych (odpowiedzi, płeć osoby ankietowanej)
Zestaw pytań ankiety wykorzystywanej do klasyfikacji płci

1. Ulubiony smak: 1-gorzki, 2-słony, 3-kwaśny, 4-słodki
2. Zależy ci aby wywrzeć dobre wrażenie na otoczeniu, liczysz się z jego opinią: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
3. Która z postaw najlepiej opisuje twój charakter: 1-typowy ekstrawertyk potrzebujący towarzystwa innych ludzi, zmian, ciągłego ruchu itp., 2-typowy introwertyk, często unika ludzi jest skupiony na swoich przeżyciach, 3-preferuje postawę otwartości wobec innych ale nie do końca, lubi ruch i zmiany ale czasami woli samotność, 4-preferuje postawę zamkniętą wobec innych, otwiera się dla niewielu wybranych, zmiany i ruch owszem ale dość rzadko.
4. Twoje decyzje są niezależne, sugerujesz się najczęściej własną wiedzą i intuicją niż radami innych: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
5. Czy inni uważają ciebie za osobę ambitną?: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
6. Czy jesteś wytrwały i konsekwentny w działaniu, osiągasz cele dzięki tym cechom?: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
7. Czy uważasz, że jesteś dokładny, precyzyjny w działaniu, zwracasz uwagę na szczegóły, drobne detale?: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
8. Wybierz jedną z odpowiedzi: 1-jesteś zdecydowanie osobą wrażliwą, świat odbierasz zmysłami, często brak ci praktycznego stosunku do życia, 2-jesteś raczej wrażliwy, świat w dużym stopniu odbierasz zmysłami, czasem brak ci praktycznego stosunku do życia, 3-jesteś raczej osobą praktyczną i trzeźwo myślącą, niezbyt często odbierasz świat emocjonalnie, masz raczej racjonalny i praktyczny stosunek do życia, 4-jesteś osobą o prawie wyłącznie praktycznym stosunku do życia, emocje uważasz najczęściej za rzecz szkodliwą
9. Decyzje podejmujesz szybko, ufasz swojej intuicji, nie lubisz zbyt długo zastanawiać się przed podjęciem decyzji: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
10. Wybierz najbliższą ci postawę: 1-zdecydowanie wolisz dominować nad otoczeniem, 2-bliższa jest ci postawa dominująca w stosunkach z innymi niż podporządkowanie się, 3- w zasadzie możesz być stroną dominującą lub podporządkować się, nie sprawia ci to żadnej różnicy, 4-wolisz być raczej w cieniu, nie masz skłonności do dominacji.
11. Uważasz, że jesteś osobą mówiącą wprost, czasem nawet nietaktownie, często nie zdając sobie z tego sprawy: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
12. Spór należy załagodzić i znaleźć z niego inne wyjście niż np. otwarta walka lub zemsta: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
13. Czy często jesteś zgryźliwy i złośliwy, lubisz krytykować innych nawet bez wyraźnej potrzeby: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
14. Jesteś osobą upartą i nieprzejednaną w swoich poglądach i postawach: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
15. Jesteś osobą nie pozbawioną, łatwo ciebie urazić niezbyt miłym dla ciebie stwierdzeniem: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
16. Czy jesteś osobą tolerancyjną, pochwalasz różne nietypowe zachowania, postawy: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
17. Czy jesteś optymistą w większości spraw: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
18. Czy inni uważają, że jesteś osobą pracowitą: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
19. Czy jesteś osobą oszczędną: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
20. Czy uważany jesteś za osobę energiczną: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
21. Czy można na tobie polegać: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
22. Czy jesteś osobą punktualną: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
23. Czy jesteś osobą wesołą i pogodną: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
24. Czy pracując bądź też ucząc się wolisz to robić we współpracy z innymi: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie
25. Czy jesteś osobą gadatliwą: 1-zdecydowanie tak, 2-raczej tak, 3-raczej nie, 4-zdecydowanie nie

Wprowadzenie

Wykorzystanie sieci neuronowych w psychologii jest znane od dawna. Potencjał zdolności do klasyfikacji wykorzystywany jest zarówno przy wykorzystaniu sieci neuronów uczonych pod nadzorem (znane są odpowiedzi) jak również w przypadku sieci neuronowych uczonych bez nadzoru „samoorganizujących się” (odpowiedzi nie są znane – następuje na wyjściu sieci neuronowej podział danych wejściowych na kategorie). Przedmiotem nauki sieci neuronowej może być np. profil psychologiczny osoby na podstawie ankiety z pytaniami. Na wejście sieci neuronowej podawać można odpowiedzi na ankietę psychologiczną a na wyjściu sieci neuronowej cel analizy czyli wybrana cecha (np. ekstrawertyzm lub introwertyzm osoby ankietowanej)

Istotnym celem pracy psychologów jest m.in. tworzenie takich testów, które w możliwie najprostszej i najkrótszej formie pozwolą zmierzyć w najbardziej rzetelny sposób wybraną cechę. W ramach pracy dyplomowej [3] Kazimierz Roguski przeanalizował tzw. Test Postaw Miłosnych autorstwa C. i S. Hendrick. Celem badań było zmniejszenie ogólnej liczby pozycji testowych. Odpowiedni wybór spośród pytań testowych dla każdej z sześciu skal umożliwił otrzymanie rzetelnego wyniku nie na podstawie siedmiu, ale trzech lub co najwyżej czterech z nich. Całość badań wykonana została przy wykorzystaniu sieci neuronowej wielowarstwowej w programie *Neuronix v.3.0*.

Celem ćwiczenia jest nauczanie sieci neuronowej klasyfikacji płci ankietowanych osób na podstawie odpowiedzi na ankietę. Wykorzystanych zostanie ponad 600 kompletnych ankiet do rozwiązania problemu rozpoznawania płci.

Cel ćwiczenia

Celem ćwiczenia jest poznanie możliwości wykorzystania sieci neuronowych typu MLP i RBF do zadania klasyfikacji (uczenie nadzorowane).

Dodatkowo zostanie wykonana analiza grupowania danych przez sieć Kohonena (uczenie nie nadzorowane) – zweryfikowane zostanie czy podział na 2 grupy jest podziałem wg płci.

Analiza danych wykonana zostanie w arkuszu kalkulacyjnym – zbieżność odpowiedzi na dane pytanie u kobiet i mężczyzn (współczynnik korelacji liniowej Pearsona) oraz średnia wartość odpowiedzi u kobiet i mężczyzn. Oba parametry mogą być pomocne w wyborze właściwych wejść sieci neuronowej (pytania różnicujące płeć).

Opis analizy statystycznej w arkuszu excel:

Pierwszym zadaniem jest redukcja liczby wejść sieci neuronowej. Wykorzystujemy do tego celu plik **statystyka_plec.xls**. Należy policzyć procentowy rozkład odpowiedzi na dane pytanie dla poszczególnych płci czyli wypełnić formułami wiersze **K1** (procent wyboru pozycji 1 odpowiedzi kobiet na danym pytaniu) do **K4** (procent wyboru pozycji 4 odpowiedzi kobiet na danym pytaniu) oraz **M1** do **M4** wg analogicznej reguły. Odpowiedzi ankiety zostały posortowane. Wiersze od 16 do 294 zawierają odpowiedzi kobiet, natomiast wiersze od 295 do 579 zawierają odpowiedzi mężczyzn. Wykorzystać można następującą formułę dla komórki B2 (procent wyboru pozycji 1 odpowiedzi kobiet na pytaniu nr.1):

=LICZ.JEŻELI(B16:B294;1)/ILE.LICZB(B16:B294)*100. Funkcja LICZ.JEŻELI oblicza liczbę komórek z zakresu B16:B294, których wartość wynosi 1. Funkcja ILE.LICZB oblicza liczbę komórek z podanego zakresu zawierających liczby. Formułę można skopiować na pozostałe komórki w wierszu oznaczonym jako **K1**. Analogicznie liczymy wiersz oznaczony jako **K2** (zmiana: formuła LICZ.JEŻELI liczy ilość cyfr 2 z podanego zakresu. Formuła z komórki B2 ma również zastosowanie do B6 (procent wyboru pozycji 1 odpowiedzi mężczyzn na pytaniu nr.1), przy czym zakres komórek to: B295: B579. Po wypełnieniu formułami wierszy **K1 .. K4, M1..M4** możemy policzyć współczynniki korelacji odpowiedzi kobiet i mężczyzn na dane pytanie dla wszystkich kolejnych pytań.

Współczynnik korelacji jest tym bliższy wartości 1, im dwie serie danych dla których jest liczony są bardziej zbliżone. Im niższy współczynnik korelacji tym w naszym przypadku większe różnice w wyborze odpowiedzi od 1 do 4 na danym pytaniu. Dla pytania pierwszego (komórka B10) formuła jest następująca:

=WSP.KORELACJI(B2:B5;B6:B9) gdzie zakres komórek B2:B5 to seria pierwsza danych (procent odpowiedzi kobiet na każde z 4 możliwych odpowiedzi), a zakres komórek B6:B9 to druga seria danych (procent odpowiedzi mężczyzn na każde z 4 możliwych odpowiedzi). Ostatnim elementem jest wpisanie subiektywnej oceny przydatności danego wejścia w wierszu opisanym jako **wybór wejścia**.

Zadanie klasyfikacji i grupowania wykonane zostaną w programie Statistica.

Opis wykonania ćwiczenia (sieć neuronowa MLP, RBF w programie Statistica)

w statistica: Menu -> nowy -> arkusz – liczba zmiennych 28, liczba przypadków 745

w excel zaznaczamy zakres A1;AB746 czyli z nagłówkami całość danych z zakładki "do_statistica_dane"

w statistica wklejamy do arkusza -> wklej z nagłówkami -> wklej z nazwami zmiennych

menu -> Data Mining -> Sieci neuronowe
wybieramy -> nowa analiza -> klasyfikacja

okno SANN wybór danych
zakładka "podstawowe"
wybieramy -> projekt sieci użytkownika
klikamy w pole "zmienne"
wybieramy "wyjścia jakościowe" - PLEC
wybieramy "wejścia ilościowe" - tutaj wskazujemy (trzymając naciśnięty klawisz CTRL)
wybrane przez nas zmienne objaśniające bez zmiennej "KOD"

klikamy w zakładkę "wybór podzbiorów" i zaznaczamy pole "identyfikator próby"
klikamy w pole "ucząca" - klikamy w "stan" włączona - klikamy w pole "zmienna identyfikująca próby" wybieramy zmienną "kod_danych", w polu "kod próby uczącej" wpisujemy "1", klikamy ok.
klikamy w pole "testowa" - klikamy w "stan" włączona - klikamy w pole "zmienna identyfikująca próby" wybieramy zmienną "kod_danych", w polu "kod próby testowej" wpisujemy "2", klikamy ok.

zamykamy okno "SANN wybór danych klikając pole OK.

pojawia się zakładka "SANN projekt użytkownika"

w zakładce "podstawowe"
wybieramy -> typ sieci -> perceptron wielowarstwowy (w dalszych próbach można również testować sieć RBF)
wybieramy funkcję błędu „Entropia wzajemna” lub „suma kwadratów”

Opis funkcji błędów

Suma-Kwadratów. Błąd jest sumą kwadratów różnic pomiędzy wartościami zadanymi i wartościami otrzymanymi na wyjściach każdego neuronu wyjściowego. Jest to standardowa funkcja błędu stosowana w trakcie uczenia sieci neuronowych. Jej wybór jest z pewnością najwłaściwszą decyzją w większości problemów regresyjnych. Zapewnia ona również możliwą do przyjęcia sprawność w problemach klasyfikacyjnych, chociaż można spotkać się z opiniami (często zasadnymi), że w tym ostatnim przypadku właściwsze jest stosowanie funkcji entropii.

Entropia wzajemna (pojedyncza i wielokrotna). Błąd ten jest sumą iloczynów zadanych wartości oraz logarytmów błędów dla każdego neuronu wyjściowego. Funkcja ta występuje w dwóch wersjach: jedna z nich przeznaczona jest dla sieci z pojedynczym neuronem wyjściowym (rozdzielającym dwie klasy), zaś druga dla sieci z wieloma neuronami wyjściowymi. Funkcja błędu oparta na entropii wzajemnej została specjalnie zaprojektowana dla problemów klasyfikacyjnych, gdzie jest stosowana łącznie z występującą w warstwie wyjściowej sieci logistyczną funkcją aktywacji (gdy przewidywane jest pojedyncze wyjście sieci) lub funkcją aktywacji typu Softmax (gdy w sieci występuje większa liczba wyjść).

wyberamy funkcje aktywacji dla neuronów ukrytych oraz neuronu wyjściowego

Opis funkcji aktywacji

Liniowa. Działanie programu przy wyborze funkcji liniowej polega na bezpośrednim przekazaniu wartości wyrażającej łączne pobudzenie neuronu na jego wyjście. Tego typu zasada działania neuronów stosowana jest w różnych typach sieci, w tym oczywiście w sieciach liniowych, ale występuje również w innych sieciach, na przykład w warstwie wyjściowej sieci o radialnych funkcjach bazowych.

Softmax. Jest to w istocie funkcja wykładnicza, której wartość zostaje jednak dodatkowo znormalizowana w taki sposób, aby suma aktywacji dla całej warstwy była równa 1. Funkcja ta, używana jest najczęściej w warstwie wyjściowej w perceptronie wielowarstwowym dostosowanym do rozwiązywania problemów klasyfikacyjnych. Dzięki zastosowanej normalizacji wartości wyjściowe mogą być interpretowane jako oszacowania prawdopodobieństw przynależności danego sygnału wejściowego do poszczególnych klas. Zakres (0,+1)

$$\frac{e^{x_i}}{\sum_i e^{x_i}}$$

Hiperboliczna. Funkcja aktywacji często bywa przybliżana za pomocą funkcji matematycznej tangens hiperboliczny (oznaczanej tanh, tgh). Podobnie jak w przypadku funkcji logistycznej, jest to krzywa S-kształtna, z tą różnicą, że wartości wyjściowe ulokowane są w przedziale (-1,+1). Z uwagi na symetrię funkcji tanh często działa ona lepiej niż funkcja logistyczna. Funkcja taka może się doskonale sprawdzić w licznych nieliniowych sieciach neuronowych - szczególnie w perceptronach wielowarstwowych.

$$\frac{e^x - e^{-x}}{e^x + e^{-x}}$$

wyberamy liczbę sieci oraz liczbę neuronów w warstwie ukrytej

w zakładce "perceptron"

wybieramy algorytm uczenia, liczbę epok oraz rodzaj inicjalizacji sieci

w zakładce "wykres uczenia"

wybieramy - wykres uczenia dla błędów w próbie uczącej i testowej

w oknie głównym "SANN projekt użytkownika" klikamy w przycisk "uczenie"

aby zobaczyć liczbę epok uczących należy rozszerzyć kolumnę "algorytm"

klikamy w "wybór sieci" i wskazujemy najlepszą sieć - z największą liczbą trafień w zakresie "jakość testowania" - ta miara błędu to po prostu procent trafień.

W zakładce „szczegóły” klikając w pole "globalna analiza wrażliwości" - uzyskujemy informacje o ważności poszczególnych zmiennych objaśniających w modelu prognostycznym

–
opis analizy wrażliwości w "dymku"

Fragment help programu Statistica

Globalna analiza wrażliwości. Globalna analiza wrażliwości daje pojęcie o tym, jak ważne są poszczególne zmienne wejściowe sieci. Wykonanie analizy wrażliwości polega na sprawdzeniu jak zachowuje się błąd sieci w przypadku gdy coś złego dzieje się ze zmiennymi niezależnymi. Konkretnie, po kolei dla każdej zmiennej wejściowej jej wartości zamieniane są na średnią (ze zbioru uczącego). Tak więc zmienna przestaje wносить jakąkolwiek informację. Po podaniu tak zmodyfikowanych danych na wejście sieci sprawdza się końcowy błąd predykcji. Błąd ten może poważnie wzrosnąć, albo wzrosnąć nieznacznie lub wcale. Oznacza to, że sieć jest albo bardzo wrażliwa na daną zmienną wejściową, albo też sieć na tej zmiennej zupełnie nie zależy. W arkuszu, dla każdej sieci podany jest iloraz wskazujący przyrost błędu przy usunięciu danej zmiennej wejściowej. Jeżeli wartość jest 1 lub mniejsza to sieć działa lepiej bez danej zmiennej - znak, że należy ją usunąć na stałe. Jednak pamiętać trzeba, że analiza dotyczy konkretnej sieci. Tymczasem zmienne bywają na różne sposoby powiązane, skorelowane i wykazują redundancje. Dlatego różne sieci mogą "wybrać" jako ważne różne zmienne. Dopiero wykonanie analizy wrażliwości dla wielu modeli i powtarzalność wyników powinny być podstawą do wyciągania praktycznych wniosków na temat zmiennych.

ewentualnie korygujemy później model wybierając nowy, lepszy zestaw zmiennych objaśniających

w prawym dolnym rogu okna głównego - opcja "próby" zaznaczamy pozycję "uczenie" oraz "test"

zaznaczamy pole „trafność”

klikamy w przycisk "predykcja"

kolumna PLEC wyjscie - wyjscie to klasyfikacja dla zakresu treningowego oraz testowego (na czerwono widać nietrafienia)-wklejamy ten blok danych do excel (UWAGA: jako wklej specjalnie wartości-liczby) -zakładka "wyniki_klasyfikacji siec MLP RBF" - kolumna C w excel pojawiają się wyniki: błędy MAPE, procent trafień dla zakresu treningowego oraz zakresu testowego.

W zakładce „przewidywania” zaznaczamy tylko „uczenie”

W zakładce „szczegóły” klikając w pole "macierz pomyłek" - uzyskujemy informacje o rozkładzie trafień na danych uczących dla każdej płci. Notujemy informacje w arkuszu kalkulacyjnym w wynikach.

W zakładce „przewidywania” zaznaczamy tylko „test”

W zakładce „szczegóły” klikając w pole "macierz pomyłek" - uzyskujemy informacje o rozkładzie trafień na danych testowych dla każdej płci. Notujemy informacje w arkuszu kalkulacyjnym w wynikach.

kolejne kroki: szukamy właściwych parametrów sieci dla których liczba trafień w zakresie danych testowych będzie największa-można wykorzystać narzędzie "więcej (projekt automatyczny),

Poszukiwanie właściwego modelu to również manipulowanie doбором zmiennych objaśniających na podstawie analizy wrażliwości oraz wstępnego doboru zmiennych w excel. Można też sprawdzić inne algorytmy uczące.

Opis wykonania ćwiczenia (sieć neuronowa Kohonena w programie Statistica)

w statistica: Menu -> nowy -> arkusz – liczba zmiennych 28, liczba przypadków 745

w excel zaznaczamy zakres A1;AB746 czyli z nagłówkami całość danych z zakładki "do_statistica_dane"

w statistica wklejamy do arkusza -> wklej z nagłówkami -> wklej z nazwami zmiennych

menu -> Data Mining -> Sieci neuronowe

wybieramy -> nowa analiza -> analiza skupień

Zakładka ‘podstawowe’->zmiennie ->wejścia ilościowe (od 2 do 27) czyli bez KOD oraz PLEC lub tylko te które wcześniejsza analiza uznała za najwłaściwsze.

Zakładka „wybór podzbiorów”->identyfikator próby->

klikamy w pole "ucząca" - klikamy w "stan" włączona - klikamy w pole "zmienna identyfikująca próby" wybieramy zmienną "kod_danych", w polu "kod próby uczącej" wpisujemy "1", klikamy ok.

klikamy w pole "testowa" - klikamy w "stan" włączona - klikamy w pole "zmienna identyfikująca próby" wybieramy zmienną "kod_danych", w polu "kod próby testowej" wpisujemy "2", klikamy ok.

zamykamy okno "SANN wybór danych klikając pole OK.

pojawia się zakładka "SANN projekt użytkownika"

w zakładce "podstawowe Kohonen" podajemy wysokość 1, szerokość 2 (oznacza to, że wykorzystamy 2 neurony wyjściowe do grupowania danych)

zakładka „uczenie Kohonena” -> wybieramy początek 1 (początkowy promień sąsiedztwa, koniec 0 (końcowy promień sąsiedztwa na końcu procesu uczenia).

Na karcie Uczenie Kohonena okna SANN - Projekt użytkownika dostępne są opisane poniżej opcje.

Uczących. W panelu tym można określić opcje dla algorytmu uczącego Kohonena stosowanego w analizie skupień.

Liczba epok. W polu tym należy wpisać liczbę epok uczenia.

Szybkość uczenia. Szybkość uczenia Kohonena zmienia się liniowo od pierwszej do ostatniej epoki uczenia. Można określić wartość Początkową i Końcową.

Sąsiedztwo. Parametr ten to "promień" sąsiedztwa (kwadratowego) danego neuronu. Na przykład dla sąsiedztwa = 2, sąsiedzi tworzą kwadrat 5x5.

Jeżeli dany, zwycięski neuron znajduje się przy brzegu mapy, to jego sąsiedztwo kończy się na brzegu.

Promień sąsiedztwa zmieniany jest liniowo od wartości **początkowej** do **końcowej**.

Warunki zatrzymania. Opcje dostępne w tym polu umożliwiają określenie kiedy należy zastosować warunki dla dostatecznie wczesnego zatrzymania uczenia sieci. Uwaga. Jeżeli z danych nie wydzielono próby testowej to SANN kończy uczenie sieci neuronowej na podstawie danych z próby uczącej. Jednakże da się zastosować dane uczące do określania, kiedy zakończyć uczenie również jeżeli wybierzemy część danych, które będą służyć wyłącznie do oceny trafności modelu (tj. nie są wykorzystywane przy uczeniu sieci). Aby zatrzymać uczenie na podstawie danych uczących przy wydzielonym podzbiorze do oceny modelu, oznaczamy dane do oceny jako próbę walidacyjną a nie testową. Zwróćmy uwagę, że różnica między próbami polega na tym, że próba testowa jest używana podczas uczeni do określania, kiedy je zatrzymać, a próba walidacyjna w ogóle nie jest używana w uczeniu sieci. Dzięki wskazaniu próby walidacyjnej przy niewybranej próbie testowej uzyskujemy zatrzymywanie uczenia w oparciu o zbiór uczący i możliwość oceny trafności sieci na danych nieużywanych przy uczeniu.

Zastosuj. Zaznaczenie tego pola wyboru umożliwia dostatecznie wczesne zatrzymanie uczenia sieci neuronowej. Wczesne zatrzymanie stosowane jest, gdy algorytm uczący spełni zdefiniowane poniżej warunki.

Poprawa. Gdy stosowane są warunki zatrzymania, uczenie sieci zakończy się jeśli średnia poprawa błędu sieci w ciągu określonej liczby epok będzie mniejsza od wartości podanej w polu Poprawa.

Liczba epok. W polu tym podajemy liczbę epok uczenia w ciągu których średnia poprawa błędu sieci musi być przynajmniej tak duża jak wartość określona w polu Poprawa (patrz powyżej).

Inicjalizacja sieci. Opcje dostępne w tym polu umożliwiają określenie sposobu inicjalizacji wag (czyli ustawienie ich wartości) na początku uczenia. Można wybrać Losowa, gaussowska lub Losowa, równomierna. Oprócz wyboru rozkładu, należy również określić wartość średnia/min oraz wariancja/maks. Można dowolnie zmieniać domyślne ustawienia tych parametrów.

Losowa, gaussowska. Wybranie tej opcji spowoduje losową, gaussowską inicjalizację wag dla modelu sieci neuronowych do określenia początkowych wartości wag. Wykorzystany zostanie rozkład normalny (o średniej i wariancji określonej poniżej).

Losowa, równomierna. Wybranie tej opcji spowoduje losową, równomierną inicjalizację wag dla modelu sieci neuronowych do określenia początkowych wartości wag. Wykorzystany zostanie rozkład równomierny (o wartości minimalnej i maksymalnej określonej poniżej).

Średnia/Min. W polu tym należy określić wartość średnią (dla rozkładu normalnego) lub wartość minimalną (dla rozkładu równomiernego) generatora liczb losowych, który określi początkowe (tzn. przed rozpoczęciem uczenia) wagi sieci.

Wariancja/Maks. W polu tym należy określić wariancję (dla rozkładu normalnego) lub wartość maksymalną (dla rozkładu równomiernego) generatora liczb losowych, który określi początkowe (tzn. przed rozpoczęciem uczenia) wagi sieci.

Zakładka „wykres uczenia”-wybieramy wykres błędu dla próby uczącej (na niej sieć neuronowa ustala wartości wag w wyniku samoorganizacji) i testowej.

Klikamy w przycisk „uczenie”

Po zakończeniu nauki;

Klikamy w Przyciski „liczności – jeden poda rozkład słupkowy, drugi liczbowy liczby zestawów w każdym koszyku.

Klikamy w zakładkę „wykres Kohonena”- widać rozkład zestawów (skupienie w obu koszykach)

Klikamy w zakładkę „podsumowanie” – widać zbiorcze parametry uczenia.

Kopiujemy wyniki do arkusza kalkulacyjnego

W programie Statistica

Zakładka „predykcja”-zaznaczamy pole „neuron zwycięski”

Zaznaczamy próba pole „test” aby otrzymać wyniki dla zakresu uczenia i testowania

Klikamy w przycisk „predykcja”

Kopiujemy kolumnę „pozycja” do arkusza kalkulacyjnego do zakładki „wyniki sieć Kohonena”, kolumna C

Otrzymaliśmy przypisanie każdego zestawu uczącego do “koszyka czyli danego neuronu wyjściowego” (1,2) lub (1,1)

Dla każdego koszyka możemy sprawdzić jego zawartość tzn. wykonać analizę ile zestawów ankietowanych mężczyzn oraz kobiet jest w każdym koszyku. Na tej podstawie możemy podać wniosek (lub nie), że dany koszyk odpowiada za zgrupowanie ankiet danej płci.

Element dodatkowy

Możemy postawić hipotezę – powinny być 3 grupy – mężczyźni, kobiety oraz osoby, których odpowiedzi nie były zgodne z podaną płcią. Wykonujemy te same czynności ale przy podaniu wysokość 1, szerokość 3. Uwaga: koszyki/neurony najbardziej oddalone reprezentują najbardziej odmienne grupy.

Otrzymaliśmy przypisanie każdego zestawu uczącego do “koszyka czyli danego neuronu wyjściowego” (1,2), (1,3) lub (1,1)

Dla każdego koszyka możemy sprawdzić jego zawartość tzn. wykonać analizę ile zestawów ankietowanych mężczyzn oraz kobiet jest w każdym koszyku. Na tej podstawie możemy podać wniosek (lub nie), że dany koszyk odpowiada za zgrupowanie ankiet danej płci oraz trzeciej grupy.

Sprawozdanie: 1 wstęp, 2. Tabelaryczna i graficzna prezentacja wyników, 3. Wnioski