

IBM SPSS Neural Networks 29



Uwaga

Przed użyciem tych informacji i produktu, którego one dotyczą, przeczytaj informacje znajdujące się w sekcji “Uwagi” na stronie 17.

Informacje o produkcie

Niniejsze wydanie dotyczy wersji 29, wydania 0, modyfikacji 1 produktu IBM® SPSS Statistics oraz wszystkich kolejnych wydań i modyfikacji, dopóki nie zostanie to określone inaczej w nowych wydaniach.

© Copyright International Business Machines Corporation .

Spis treści

Rozdział 1. Sieci neuronowe.....	1
Wprowadzenie do sieci neuronowych.....	1
Czym jest sieć neuronowa?.....	1
Struktura sieci neuronowej.....	2
Perceptron wielowarstwowy (MLP).....	3
Podzbiory	5
Architektura	6
Uczenie	7
Dane wyjściowe	8
Zapisywanie	10
Eksport	10
Opcje	10
radialna funkcja bazowa.....	11
Podzbiory	13
Architektura	14
Dane wyjściowe	14
Zapisywanie	15
Eksport	16
Opcje	16
Uwagi.....	17
Znaki towarowe.....	18
Indeks.....	21

Rozdział 1. Sieci neuronowe

Następujące funkcje dotyczące sieci neuronowych są dostępne w SPSS Statistics edycji Premium Edition lub opcjonalnym module Neural Networks.

Wprowadzenie do sieci neuronowych

Sieci neuronowe, ze względu na swoje możliwości, elastyczność i łatwość stosowania, są preferowanym narzędziem w wielu zastosowaniach związanych z predykcyjną eksploracją danych. Predykcyjne sieci neuronowe są szczególnie przydatne w zastosowaniach, w których bazowy proces jest złożony, takich jak:

- Prognozowanie zapotrzebowań klientów upraszcza koszty produkcji i dostaw.
- Prognozowanie prawdopodobieństwa reakcji na przekazy marketingowe rozsyłane pocztą, w celu ustalenia, do których gospodarstw domowych z listy mailingowej należy rozesłać ofertę.
- Ocena wnioskodawców w celu określenia ryzyka udzielenia kredytu.
- Wykrywanie oszukańczych transakcji w bazie danych o roszczeniach ubezpieczeniowych.

Sieci neuronowe wykorzystywane w zastosowaniach prognozowych, takich jak sieci perceptronu wielowarstwowego (MLP) i radialnej funkcji podstawy (RBF), są nadzorowane w tym sensie, że prognozowane wyniki modelu mogą być porównywane z znanymi wartościami zmiennych docelowych. Opcja Neural Networks umożliwia dopasowywanie sieci MLP i RFG oraz zapisywanie wyników modeli do oceniania.

Czym jest sieć neuronowa?

Terminu **sieć neuronowa** używa się na określenie całej rodziny luźno powiązanych modeli, których cechą charakterystyczną jest duża przestrzeń parametrów i elastyczna struktura. Modele te opracowano, wzorując się na działaniu mózgu. Nowe modele sieci neuronowych przeznaczone były w większości do zastosowań niezwiązanych z biologią, jednak stosowana terminologia odzwierciedla pierwotne inspiracje neurologiczne.

W różnych dziedzinach stosuje się różne definicje sieci neuronowych. Chociaż żadna pojedyncza definicja nie obejmuje właściwie całej rodziny modeli, należy wziąć pod uwagę opis w publikacji¹:

Sieć neuronowa to masowo równoległy procesor rozproszony, który jest naturalnie predystynowany do gromadzenia wiedzy eksperymentalnej i udostępniania jej w celu wykorzystania. Sieć neuronowa pod dwoma względami przypomina mózg:

- Sieć zdobywa wiedzę, ucząc się.
- Do przechowywania wiedzy wykorzystywane są siły połączeń między neuronami, nazywane wagami synaptycznymi.

Dyskusję na temat tego, dlaczego ta definicja jest być może zbyt restrykcyjna, zawiera publikacja².

Dla odróżnienia sieci neuronowych od tradycyjnych metod statystycznych równie ważne, jak sam tekst definicji, jest to, czy ta definicja *nie* zawiera. Na przykład tradycyjny model regresji liniowej może uzyskiwać wiedzę metodą najmniejszych kwadratów i przechowywać ją w postaci współczynników regresji. W tym sensie jest siecią neuronową. Na tej podstawie można byłoby twierdzić, że regresja liniowa jest szczególnym przypadkiem pewnej kategorii sieci neuronowych. Jednak regresja liniowa jest modelem o sztywnej strukturze, w którym pewne założenia narzucone są jeszcze przed rozpoczęciem uczenia na podstawie danych.

¹ Haykin, S. 1998. *Neural Networks: A Comprehensive Foundation*, 2nd ed. New York: Macmillan College Publishing.

² Ripley, B. D. 1996. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

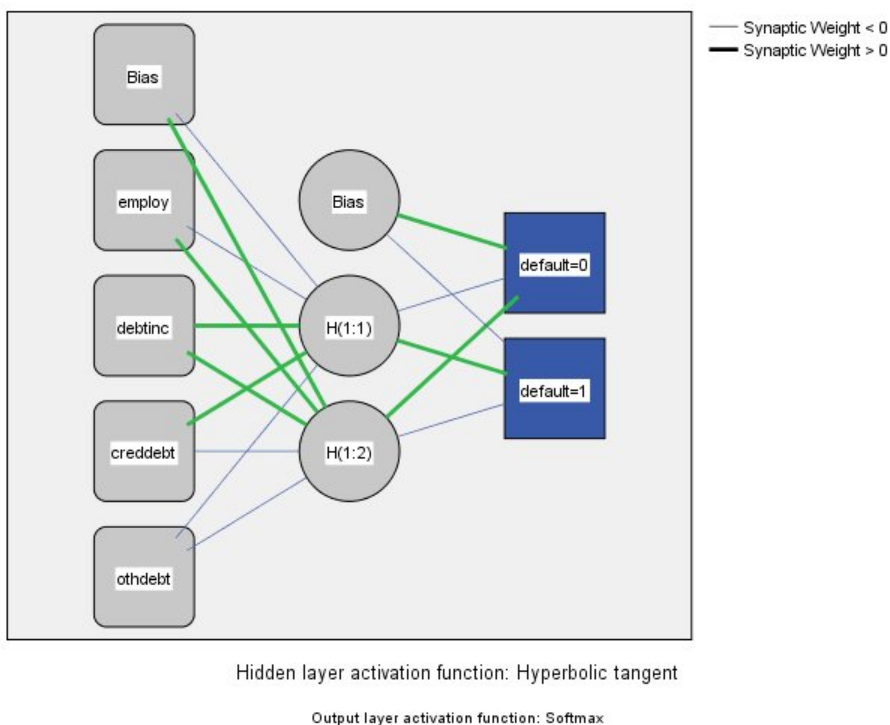
Natomiast przedstawiona powyżej definicja stawia tylko minimalne wymagania co do struktury modelu i dotyczących go założeń. Zatem sieć neuronowa może być przybliżeniem wielu różnych modeli statystycznych, nie wymagając wcześniejszego stawiania hipotez co do relacji między zmiennymi zależnymi i niezależnymi. Postać relacji wyznaczana jest w procesie uczenia sieci. Jeśli odpowiednia jest liniowa zależność między zmienną zależną i niezależną, wyniki działania sieci neuronowej powinny być dobrym przybliżeniem modelu regresji liniowej. Jeśli bardziej odpowiednie jest powiązanie nieliniowe, wówczas sieć neuronowa będzie automatycznie przybliżać „prawidłową” strukturę modelu.

Z tą elastycznością wiąże się jednak pewien koszt: trudność w interpretacji wag synaptycznych sieci neuronowej. Dlatego, jeśli celem analityka jest wyjaśnienie procesu, który prowadzi do powstania relacji między zmienną zależną i niezależną, lepiej byłoby użyć bardziej tradycyjnego modelu statystycznego. Jeśli jednak łatwość interpretacji modelu nie jest ważna, w wielu przypadkach sieć neuronowa pozwala na szybsze uzyskanie dobrych wyników.

Struktura sieci neuronowej

Mimo że sieci neuronowe nakładają tylko minimalne wymagania na strukturę i założenia modelu, przydatne jest zrozumienie ogólnej **architektury sieci**. Sieć typu Perceptron wielowarstwowy (MLP) lub Radialna funkcja bazowa (RBF) to funkcja predyktorów (nazywanych także wejściami lub zmiennymi niezależnymi), która minimalizuje błąd predykcji zmiennych przewidywanych (nazywanych także wyjściami).

Weźmy pod uwagę zbiór danych *bankloan.sav* dostarczony z produktem i założmy, że chcemy rozpoznać potencjalnych klientów niespłacających kredytów w puli osób składających wnioski kredytowe. Sieć MLP lub RBF zastosowana do rozwiązania tego problemu jest funkcją pomiarów, która minimalizuje błąd w predykcji niespłacenia kredytu. Poniższy rysunek ułatwi zrozumienie postaci tej funkcji.



Rysunek 1. Architektura skierowana (feedforward) z jedną warstwą ukrytą

Ta struktura nazywana jest **architekturą skierowaną**, ponieważ połączenia w sieci są skierowane od warstwy wejściowej do warstwy wyjściowej bez żadnych pętli zwrotnych. Na tym rysunku:

- **Warstwa wejściowa** zawiera predyktory.
- **Warstwa ukryta** zawiera węzły nieobserwowalne (jednostki). Wartość każdej ukrytej jednostki to pewna funkcja predyktorów; dokładna forma funkcji zależy częściowo od typu sieci i częściowo od specyfikacji kontrolowanych przez użytkownika.

- **Warstwa wyjściowa** zawiera odpowiedzi. Ponieważ informacja o tym, czy ktoś w przeszłości nie spłacał w terminie kredytu, jest zmienną jakościową dwóch kategorii, rekodowana jest do postaci dwóch zmiennych wskaźnikowych. Każda jednostka wyjściowa jest pewną funkcją jednostek ukrytych. Również tutaj dokładna postać funkcji zależy po części od typu sieci, a po części od ustawień określonych przez użytkownika.

Sieć MLP pozwala na stosowanie drugiej warstwy ukrytej; w takim przypadku każda jednostka drugiej warstwy ukrytej jest funkcją jednostek pierwszej warstwy ukrytej, a każda odpowiedź jest funkcją jednostek drugiej warstwy ukrytej.

Perceptron wielowarstwowy (MLP)

Procedura Perceptron wielowarstwowy (MLP) tworzy model predykcyjny dla co najmniej jednej zmiennej zależnej (wynikowej) w oparciu o wartości zmiennych predykcyjnych.

Przykłady. Poniżej przedstawiono dwa scenariusze wykorzystujące procedurę MLP:

Specjalista ds. kredytów w banku chce zidentyfikować cechy, które są charakterystyczne dla osób, które zalegają ze spłatą pożyczek, a następnie chce użyć tych cech do określenia dobrego i złego ryzyka kredytowego. Korzystając z próby składającej się z klientów z przeszłości, może wyuczyć procedurę perceptronu wielowarstwowego, sprawdzić poprawność analizy z użyciem próby wstrzymania klientów z przeszłości, a następnie użyć sieci do sklasyfikowania przyszłych klientów pod kątem dobrego lub złego ryzyka kredytowego.

System szpitalny jest zainteresowany śledzeniem kosztów i długości pobytu przyjętych w celu leczenia zawału mięśnia sercowego (MI lub atak serca). Uzyskanie dokładnych oszacowań tych miar umożliwi administracji prawidłowe zarządzanie dostępnymi łóżkami w trakcie leczenia pacjentów. Dzięki użyciu rekordów leczenia pacjentów tworzących próbę, którzy byli poddani leczeniu MI, administrator może wyuczyć sieć, tak aby przewidywała koszty i długość pobytu.

Zagadnienia dotyczące danych

Zmienne zależne. Możliwe zmienne zależne to:

- *Nominalny.* Zmienna może być traktowana jako nominalna, gdy jej wartości reprezentują kategorie bez wewnętrznego rankingu (na przykład dział przedsiębiorstwa, w którym pracuje pracownik). Przykładami zmiennych nominalnych są: region, kod pocztowy lub wyznanie.
- *Porządkowy.* Zmienna może być traktowana jako porządkowa, gdy jej wartości reprezentują kategorie z jakimś nieodłącznym rangą (na przykład poziomy zadowolenia z usługi z bardzo niezadowolonego do bardzo zadowolonego). Przykładami zmiennych porządkowych mogą być oceny opinii reprezentujące stopień satysfakcji lub przekonania oraz oceny preferencji.
- *scale.* Zmienna może być traktowana jako skala (ilościowa), gdy jej wartości reprezentują uporządkowane kategorie z miarodajnym pomiarem, tak aby porównania odległości między wartościami były odpowiednie. Przykładami zmiennych ilościowych mogą być: wiek w latach lub przychód w tysiącach złotych.

W procedurze przyjmuje się, że odpowiedni poziom pomiaru został przypisany do wszystkich zmiennych zależnych. Można jednak tymczasowo zmienić poziom pomiaru dla zmiennej, klikając prawym przyciskiem myszy zmienną na liście zmiennych źródłowych i wybierając poziom pomiaru z menu kontekstowego.

Ikona obok każdej zmiennej na liście zmiennych określa poziom pomiaru oraz typ danych:












Tabela 1. Ikony poziomu pomiaru				
	Liczbowy	Łańcuch	Data	Czas
Zmienna (ilościowa)		nie dotyczy		

Tabela 1. Ikony poziomu pomiaru (kontynuacja)				
	Liczbowy	Łańcuch	Data	Czas
Porządkowy				
Nominalna				

Zmienne predykcyjne. Predyktory można określić jako czynniki (jakościowe) lub współzmiennie (skala).

Kodowanie zmiennych jakościowych. Procedura tymczasowo przekodowuje predyktory jakościowe i zmienne zależne, używając jednego z c kodowań na czas trwania procedury. Jeśli istnieje c kategorii zmiennej, to zmienna jest przechowywana jako c wektorów, gdzie pierwszą kategorię oznacza $(1,0,\dots,0)$, następną $(0,1,0,\dots,0)$, ..., a ostatnią $(0,0,\dots,0,1)$.

Ten schemat kodowania zwiększa liczbę wag synaptycznych i może spowodować spowolnienie uczenia; jednak zastosowanie mniej złożonych metod kodowania zwykle prowadzi do słabszego dopasowania sieci neuronowych. Jeśli uczenie sieci przebiega bardzo wolno, spróbuj zmniejszyć liczbę kategorii w predyktorach kategorialnych, łącząc podobne kategorie lub usuwając przypadki, które mają bardzo rzadkie kategorie.

Kodowanie jeden z c jest oparte na danych szkoleniowych, nawet jeśli zdefiniowano próbę testowania lub wstrzymania (patrz "Podzbiory" na stronie 5). Dlatego jeśli próba testowania lub wstrzymania zawiera obserwacje z kategoriami predyktorów, które nie są obecne w danych uczących, wówczas obserwacje te nie są używane przez procedurę ani oceniane. Jeśli próba testowania lub wstrzymania zawiera obserwacje z kategoriami zmiennych zależnych, które nie są obecne w danych uczących, wówczas obserwacje te nie są używane przez procedurę, ale mogą być oceniane.

Przeskalowanie. Ilościowe zmienne zależne i współzmiennie zostają domyślnie przeskalowane, aby usprawnić uczenie sieci. Przeskalowanie odbywa się na podstawie danych uczących, nawet jeśli zdefiniowana została próba testowania lub wstrzymania (patrz "Podzbiory" na stronie 5). Oznacza to, że w zależności od typu przeskalowania średnia, odchylenie standardowe, wartość minimalna lub wartość maksymalna współzmiennnej i zmiennej zależnej są obliczane z użyciem wyłączenie danych uczących. Jeżeli zostanie określona zmienna definiująca podziały, ważne jest, aby współzmiennie i zmienne zależne miały podobny rozkład w próbach uczących, testowania i wstrzymania.

Wagi liczebności. Wagi częstotliwości są ignorowane przez tę procedurę.

Replikacja wyników. Aby przeprowadzić dokładną replikację wyników, należy użyć tej samej wartości inicjalizacji dla generatora liczb losowych, tej samej kolejności danych oraz tej samej kolejności zmiennych oraz użyć tych samych ustawień procedury. Aby uzyskać więcej szczegółów na temat tego problemu, należy zapoznać się z tematami:

- **Generowanie liczb losowych.** Procedura ta używa generatora liczb losowych podczas losowego przypisywania podziałów, dodatkowego losowania podczas inicjalizacji wag synaptycznych, losowego dodatkowego losowania na potrzeby automatycznego wyboru architektury oraz algorytmu symulowanego wyżarzenia do inicjalizacji wagi i automatycznego wyboru architektury. Aby ponownie utworzyć takie same losowe wyniki w przyszłości, należy użyć tej samej wartości inicjalizacji dla generatora liczb losowych przed każdym uruchomieniem procedury wielowarstwowego perceptronu.
- **Kolejność obserwacji.** Metody uczenia online i z wykorzystaniem podzbiorów (patrz "Uczenie" na stronie 7) są bezpośrednio zależne od kolejności obserwacji; również uczenie całym zbiorem jest zależne od kolejności obserwacji, ponieważ inicjalizacja wag synaptycznych obejmuje dodatkowe losowanie z zestawu danych.

Aby zminimalizować wpływ kolejności, należy losowo ustawić obserwacje. Aby sprawdzić stabilność danego rozwiązania, można uzyskać kilka różnych rozwiązań z obserwacjami posortowanymi w różnej, losowej kolejności. W sytuacji, gdy wielkość pliku jest ekstremalnie duża, wykonywanych jest wiele uruchomień z użyciem próby obserwacji posortowanych w różnej kolejności losowej.

- **Kolejność zmiennych.** Na wyniki może wpływać kolejność zmiennych na liście czynników i współzmiennych, ponieważ wzór wartości początkowych przypisywanych po zmianie kolejności zmiennych jest różny. Podobnie jak w przypadku wpływu kolejności obserwacji, można wypróbować różne kolejności zmiennych (wystarczy przeciągnąć i upuścić na liście czynników i współzmiennych), aby ocenić stabilność danego rozwiązania.

Tworzenie sieci wielowarstwowego perceptronu

Z menu wybierz:

Analyze > Sieci neuronowe > Wielowarstwowy perceptron...

1. Wybierz co najmniej jedną zmienną zależną.
2. Wybierz co najmniej jeden czynnik lub współzmienną.

Opcjonalnie, na karcie Zmienne można zmienić metodę przeskalowania współzmiennych. Dostępne są następujące opcje:

- **Standaryzowane.** Odjęcie średniej i podział przez standardowe odchylenie, $(x - \text{średnia})/s$.
- **Znormalizowane.** Odjęcie wartości minimalnej i podział przez zakres, $(x - \text{min.})/(\text{maks.} - \text{min.})$. Znormalizowane wartości należą do przedziału od 0 do 1.
- **Skorygowane znormalizowane.** Skorygowana wersja odejmowania wartości minimalnej i podziału przez zakres, $[2 * (x - \text{min.}) / (\text{maks.} - \text{min.})] - 1$. Skorygowane wartości znormalizowane zawierają się w zakresie od -1 do 1.
- **Brak.** Bez przeskalowania współzmiennych.

Zmienne z nieznanym poziomem pomiaru

Alert poziomu pomiaru wyświetla się, gdy poziom pomiaru dla jednej lub większej ilości zmiennych w zbiorze danych jest nieznanany. Ponieważ poziom pomiaru wpływa na wyliczenie wyników dla tej procedury, wszystkie zmienne muszą mieć zdefiniowany poziom pomiaru.

Skanowanie danych. Odczytuje dane w aktywnym zbiorze danych i przypisuje domyślny poziom pomiaru do wszystkich zmiennych, które mają aktualnie nieznanany poziom pomiaru. Jeśli zbiór danych jest duży, może to zająć trochę czasu.

Przypisz ręcznie. Otwiera okno dialogowe, które zestawia wszystkie zmienne z nieznanym poziomem pomiaru. Można użyć tego okna dialogowego do przypisania poziomu pomiaru do tych zmiennych. Można również przypisać poziom pomiaru w Widoku zmiennych Edytora danych.

Ponieważ poziom pomiaru jest ważny dla tej procedury, nie można wejść do tego okna dialogowego w celu uruchomienia tej procedury, dopóki wszystkie zmienne nie będą miały zdefiniowanego poziomu pomiaru.

Podzbiory

Podzbiór danych. Ta grupa określa metodę podziału aktywnego zbioru danych na próby uczące, testu i wstrzymane. **Próba ucząca** składa się z rekordów danych używanych do szkolenia sieci neuronowych. Pewna wartość procentowa obserwacji w zbiorze danych musi być przypisana do próby uczącej w celu uzyskania modelu. **Próba testu** to niezależny zbiór danych używanych do nadzorowania błędów w czasie uczenia. Pozwala to zapobiec nadmiernemu wyuczeniu. Zdecydowanie zaleca się utworzenie próby testowej, a szkolenie sieci będzie bardziej wydajne, jeśli próba testu będzie mniejsza niż próba testująca. **Próba wstrzymana** jest innym niezależnym zestawem rekordów danych używanych do oceny końcowej sieci neuronowych. Błąd w próbie wstrzymanej daje „uczciwą” ocenę możliwości predykcyjnej modelu, ponieważ obserwacje wstrzymane nie zostały użyte do budowy modelu.

- **Przypisz obserwacje losowo na bazie względnej liczby obserwacji.** Należy określić względną liczbę (stosunek) obserwacji losowo przypisanych do każdej próby (uczącej, testu i wstrzymanej). W kolumnie % wskazany jest odsetek obserwacji, które zostaną przypisane do każdej próby w oparciu o podane liczby względne.

Na przykład, podanie 7, 3, 0 jako względnych liczb dla obserwacji w próbie uczącej, testu i wstrzymanej odpowiada odsetkom 70%, 30% i 0%. Podanie 2, 1, 1 jako względnych liczb odpowiada 50%, 25%

i 25%; 1, 1, 1 powoduje podział zbioru danych na trzy równe części między próbę uczącą, testu i wstrzymaną.

- **Użyj zmiennej dzielącej w celu przydzielenia obserwacji.** Określ zmienną numeryczną, która przydziela każdą obserwację w aktywnym zbiorze danych do próby uczącej, testu lub wstrzymanej. Obserwacje z wartością pozytywną są przydzielane do próby uczącej, a przypadki z wartością 0 lub mniejszą – do próby wstrzymanej. Obserwacje z systemowymi brakami danych zostają wykluczone z analizy. Wartości zdefiniowanych przez użytkownika braków danych zmiennej dzielącej zawsze są traktowane jako prawidłowe.

Uwaga: Użycie zmiennej dzielącej na podzbiory nie zagwarantuje uzyskania identycznych wyników przy kolejnych wykonaniach procedury. Więcej informacji zawiera punkt „Replikacja wyników” w głównym temacie [Perceptron wielowarstwowy](#).

Architektura

Karta Architektura służy do określania struktury sieci. Procedura może automatycznie wybrać „najlepszą” architekturę lub użytkownik może sam określić architekturę.

W wyniku automatycznego wyboru architektury tworzona jest sieć z jedną warstwą ukrytą. Gdy użytkownik określi minimalną i maksymalną dozwoloną liczbę jednostek w warstwie ukrytej, funkcja automatycznego wyboru architektury obliczy „najlepszą” liczbę jednostek w tej warstwie. Przy automatycznym wyborze architektury używane są domyślne funkcje aktywacji dla warstwy ukrytej i wyjściowej.

Wybór architektury przez użytkownika pozwala na szczegółowe sterowanie tworzeniem warstw ukrytych i warstwy wyjściowej. Ten sposób wyboru jest najbardziej przydatny, gdy wiadomo z góry, jaka architektura ma być stosowana lub gdy konieczne jest zoptymalizowanie wyników automatycznego wyboru architektury.

Ukryte warstwy

Ukryta warstwa zawiera nieobserwowalne węzły sieci (jednostki). Każda ukryta jednostka jest funkcją sumy ważonej danych wejściowych. Funkcja jest funkcją aktywacji, a wartości wag są uzyskiwane z algorytmu estymacji. Jeśli sieć ma drugą ukrytą warstwę, każda ukryta jednostka w drugiej warstwie jest funkcją sumy ważonej jednostek z pierwszej ukrytej warstwy. Ta sama funkcja aktywacji jest używana w obu warstwach.

Liczba ukrytych warstw. Wielowarstwowy perceptron może mieć jedną lub dwie ukryte warstwy.

Funkcja aktywacji. Funkcja aktywacji służy do łączenia sum ważonych jednostek w warstwie do wartości jednostek w następnej warstwie.

- **Tangens hiperboliczny.** Ta funkcja ma postać: $\gamma(c) = \tanh(c) = (e^c - e^{-c}) / (e^c + e^{-c})$. Jej argumentami są liczby rzeczywiste. Funkcja przekształca je do przedziału $(-1, 1)$. Gdy stosowany jest automatyczny wybór architektury, jest to funkcja aktywacji dla wszystkich jednostek w warstwach ukrytych.
- **Funkcja sigmoidalna.** Ta funkcja ma postać: $\gamma(c) = 1 / (1 + e^{-c})$. Jej argumentami są liczby rzeczywiste. Funkcja przekształca je do przedziału $(0, 1)$.

Liczba jednostek. Liczba jednostek w każdej warstwie ukrytej może być jawnie podana lub zostać automatycznie określona przez algorytm estymacji.

Warstwa wyjściowa

Warstwa wyjściowa zawiera zmienne docelowe (zależne).

Funkcja aktywacji. Funkcja aktywacji służy do łączenia sum ważonych jednostek w warstwie do wartości jednostek w następnej warstwie.

- **Tożsamość.** Ta funkcja ma postać: $\gamma(c) = c$. Jej argumentami są liczby rzeczywiste. Funkcja zwraca je bez zmian. Gdy stosowany jest automatyczny wybór architektury, jest to funkcja aktywacji dla wszystkich jednostek w warstwie wyjściowej, nawet jeśli istnieją ilościowe zmienne zależne.
- **Softmax.** Ta funkcja ma postać: $\gamma(c_k) = \exp(c_k) / \sum_j \exp(c_j)$. Jej argumentem jest wektor wartości rzeczywistych. Funkcja przekształca go w wektor, którego elementy należą do przedziału $(0, 1)$ i sumują

się do 1. Opcja Softmax jest dostępna tylko wtedy, gdy wszystkie zmienne zależne są jakościowe. Gdy stosowany jest automatyczny wybór architektury, jest to funkcja aktywacji dla jednostek w warstwie wyjściowej, jeśli wszystkie zmienne zależne są jakościowe.

- **Tangens hiperboliczny.** Ta funkcja ma postać: $\gamma(c) = \tanh(c) = (e^{-c} - e^c) / (e^{-c} + e^c)$. Jej argumentami są liczby rzeczywiste. Funkcja przekształca je do przedziału $(-1, 1)$.
- **Funkcja sigmoidalna.** Ta funkcja ma postać: $\gamma(c) = 1 / (1 + e^{-c})$. Jej argumentami są liczby rzeczywiste. Funkcja przekształca je do przedziału $(0, 1)$.

Przeskalowanie zmiennych ilościowych. Te elementy sterujące są dostępne tylko wtedy, gdy wybrano co najmniej jedną ilościową zmienną zależną.

- **Standaryzowane.** Odjęcie średniej i podział przez standardowe odchylenie, $(x - \text{średnia}) / s$.
- **Znormalizowane.** Odjęcie wartości minimalnej i podział przez zakres, $(x - \text{min.}) / (\text{maks.} - \text{min.})$. Znormalizowane wartości należą do przedziału od 0 do 1. Jest to wymagana metoda przeskalowania dla zmiennych zależnych od skali, jeśli warstwa wyjściowa korzysta z funkcji aktywacji sigmoid. Opcja poprawki definiuje małą liczbę ϵ stosowaną jako poprawka do formuły przeskalowania; ta korekta zapewnia, że wszystkie przeskalowane wartości zmiennej zależnej będą mieściły się w przedziale funkcji aktywacji. W szczególności wartości 0 i 1, które występują w nieskorygowanej formule, gdy x przyjmuje wartość minimalną i maksymalną, definiują granice przedziału funkcji sigmoidalnej, ale do niego nie należą. Skorygowana formuła to $[x - (\text{min} - \epsilon)] / [(\text{maks} + \epsilon) - (\text{min} - \epsilon)]$. Podaj liczbę nie mniejszą od 0.
- **Skorygowane znormalizowane.** Skorygowana wersja odejmowania wartości minimalnej i podziału przez zakres, $[2 * (x - \text{min.}) / (\text{maks.} - \text{min.})] - 1$. Skorygowane wartości znormalizowane zawierają się w zakresie od -1 do 1. Jest to wymagana metoda przeskalowania dla zmiennych zależnych od skali, jeśli warstwa wyjściowa korzysta z funkcji aktywacji tangensa hiperbolicznego. Opcja poprawki definiuje małą liczbę ϵ stosowaną jako poprawka do formuły przeskalowania; ta korekta zapewnia, że wszystkie przeskalowane wartości zmiennej zależnej będą mieściły się w przedziale funkcji aktywacji. W szczególności wartości -1 i 1, które występują w nieskorygowanej formule, gdy x przyjmuje wartość minimalną i maksymalną, definiują granice przedziału funkcji typu tangens hiperboliczny, ale do niego nie należą. Skorygowana formuła to $\{2 * [(x - (\text{min} - \epsilon))] / ((\text{maks} + \epsilon) - (\text{min} - \epsilon))\} - 1$. Podaj liczbę nie mniejszą od 0.
- **Brak.** Bez przeskalowania zmiennych ilościowych.

Uczenie

Karta Uczenie służy do określania sposobu uczenia sieci. To, które opcje uczenia są dostępne, zależy od typu uczenia i algorytmu optymalizacji.

Typ uczenia. Typ uczenia decyduje o tym, jak sieć przetwarza rekordy. Wybierz jeden z następujących typów uczenia:

- **Wsadowe.** Aktualizuje wagi synaptyczne dopiero po przejściu przez wszystkie rekordy uczące; a zatem uczenie całym zbiorem wykorzystuje informacje ze wszystkich rekordów w uczącym zbiorze danych. Uczenie całym zbiorem jest często preferowane, ponieważ bezpośrednio minimalizuje całkowity błąd; jednak uczenie całym zbiorem niekiedy musi wielokrotnie aktualizować wagi synaptyczne, zanim zadziała jedna z reguł zatrzymujących, zatem potrzebuje potencjalnie wielu przejść przez dane. Jest najbardziej użyteczne w przypadku „mniejszych” zbiorów danych.
- **Jednostkowo.** Aktualizuje wagi synaptyczne po przejściu przez każdy pojedynczy rekord uczący; a zatem uczenie jednostkowo wykorzystuje informacje z jednego rekordu naraz. Uczenie jednostkowo w sposób ciągły pobiera rekordy i aktualizuje wagi, dopóki nie zadziała jedna z reguł zatrzymujących. Jeśli mimo jednokrotnego użycia każdego rekordu żadna z reguł zatrzymujących nie zadziała, proces zaczyna powtórnie używać tych samych rekordów danych. Uczenie online jest lepsze niż uczenie całym zbiorem dla „większych” zbiorów danych z powiązаныmi predyktorami; tzn. jeśli istnieje wiele rekordów i wiele wejść, a ich wartości nie są niezależne od siebie, to uczenie online może szybciej uzyskać rozsądną odpowiedź niż uczenie całym zbiorem.
- **Podzbiorem.** Dzieli rekordy danych uczących na grupy mające w przybliżeniu równą wielkość, a następnie aktualizuje wagi synaptyczne po przejściu przez każdą grupę; zatem uczenie podzbiorem wykorzystuje informacje z grupy rekordów. W razie potrzeby te same grupy danych są używane

powtórnie. Uczenie podzbiorem jest podejściem pośrednim między uczeniem całym zbiorem a uczeniem jednostkowo i może stanowić najlepsze rozwiązanie w przypadku „średnich” zbiorów danych. Procedura może automatycznie określić liczbę rekordów uczących w podzbiore albo użytkownik może podać liczbę całkowitą większą niż 1 i mniejszą lub równą maksymalnej liczbie obserwacji, jaką można przechowywać w pamięci. Maksymalną liczbę obserwacji przechowywaną w pamięci można określić na karcie Opcje.

Algorytm optymalizacji. Jest to metoda używana do szacowania wag synaptycznych.

- **Skalowany sprzężony gradient.** Założenia, które uzasadniają zastosowanie metod gradientu sprzężonego, mają zastosowanie wyłącznie w przypadku typów uczenia całym zbiorem, dlatego ta metoda nie jest dostępna w przypadku uczenia online lub minizbiorów.
- **Spadek gradientu.** Metoda ta musi być używana z uczeniem online lub uczeniem minizbiorów; może być również używana z uczeniem całym zbiorem.

Opcje uczenia. Opcje uczenia pozwalają na dostrojenie algorytmu optymalizacji. Z reguły nie ma potrzeby zmiany tych ustawień, chyba że wystąpią problemy z estymowaniem odpowiedzi przez sieć.

Dostępne są następujące opcje algorytmu skalowanego sprzężonego gradientu:

- **Początkowa wartość Lambda.** Początkowa wartość lambda dla algorytmu skalowanego sprzężonego gradientu. Podaj liczbę większą niż 0 i mniejszą niż 0,000001.
- **Początkowa wartość Sigma.** Początkowa wartość sigma dla algorytmu skalowanego sprzężonego gradientu. Podaj liczbę większą niż 0 i mniejszą niż 0,0001.
- **Środek przedziału i Przesunięcie przedziału.** Środek przedziału (a_0) o przesunięcie przedziału (a) definiują przedział $[a_0 - a, a_0 + a]$, w którym wektory wag są generowane losowo przy symulowanym wyżarzaniu. Symulowane wyżarzanie służy do przelamania lokalnego minimum z intencją znalezienia minimum globalnego podczas stosowania algorytmu optymalizacji. Ta strategia jest używana przy inicjowaniu wag i automatycznym wyborze architektury. Podaj liczbę dla środka odstępu czasu i liczbę większą niż 0 dla przesunięcia odstępu czasu.

Dostępne są następujące opcje algorytmu spadku gradientu:

- **Wstępna wartość współczynnika uczenia.** Początkowa wartość współczynnika uczenia dla algorytmu spadku gradientu. Wyższy współczynnik uczenia powoduje, że sieć uczy się szybciej, potencjalnie kosztem stabilności. Podaj liczbę większą od 0.
- **Dolna granica współczynnika uczenia.** Dolna granica współczynnika uczenia dla algorytmu spadku gradientu. To ustawienie ma zastosowanie tylko do uczenia jednostkowo i podzbiorem. Podaj liczbę większą niż 0 i mniejszą niż początkowy współczynnik uczenia.
- **Moment.** Początkowa wartość momentu dla algorytmu spadku gradientu. Składnik momentu pomaga zapobiegać niestabilności wynikającej ze zbyt wysokiego współczynnika uczenia. Podaj liczbę większą od 0.
- **Redukcja współczynnika uczenia po liczbie epok.** Liczba epok (p) lub przejść przez dane próby uczącej wymagana do zmniejszenia początkowego współczynnika uczenia do dolnej granicy, gdy algorytm spadku gradientu używany jest w procesie uczenia jednostkowo lub podzbiorem. Liczba epok pozwala wpływać na współczynnik spadku współczynnika uczenia $\beta = (1/p \cdot K) \cdot \ln(\eta_0/\eta_{low})$, gdzie η_0 jest początkowym współczynnikiem uczenia, η_{low} jest dolną granicą współczynnika uczenia, a K jest łączną liczbą podzbiorów (lub rekordów uczących dla uczenia jednostkowo) w uczącym zbiorze danych. Podaj liczbę całkowitą większą od 0.

Dane wyjściowe

Struktura sieci. Wyświetla informacje podsumowujące na temat sieci neuronowej.

- **Opis.** Wyświetla informacje na temat sieci neuronowej, w tym zmienne zależne, liczbę jednostek wejściowych i wyjściowych, liczbę ukrytych warstw i jednostek oraz funkcje aktywacji.
- **Diagram.** Wyświetla diagram sieci w postaci nieedytowalnej. Należy zwrócić uwagę, że w miarę wzrostu liczby współzmiennych i poziomów czynników diagram staje się coraz trudniejszy do zinterpretowania.

- **Wagi synaptyczne.** Wyświetla oszacowania współczynników, które odzwierciedlają relację między jednostkami na danej warstwie a jednostkami na następnej warstwie. Wagi synaptyczne są oparte na próbie uczącej, nawet jeśli aktywny zbiór danych jest podzielony na dane uczące, testujące i wstrzymane. Należy zwrócić uwagę, że liczba wag synaptycznych może być duża, a te wagi co do zasady nie służą do interpretacji wyników działania sieci.

Wydajność sieci. Przedstawia wyniki służące do określania, czy model jest „dobry”. *Uwaga:* Wykresy w tej grupie są oparte na połączonej próbie uczącej i testującej lub tylko na próbie uczącej, jeśli nie ma próby testującej.

- **Podsumowanie modelu.** Przedstawia wyniki działania sieci neuronowej z podziałem na grupy danych oraz wyniki łączne, w tym błąd, błąd względny lub odsetek nieprawidłowych predykcji, regułę zatrzymującą uczenie oraz czas uczenia.

Gdy do warstwy wyjściowej stosowana jest funkcja aktywacji typu tożsamościowego, sigmoidalna lub typu tangens hiperboliczny, błąd jest błędem sumy kwadratów, Gdy do warstwy wyjściowej stosowana jest funkcja aktywacji softmax, jest to błąd entropii krzyżowej.

Prezentowane są błędy względne lub odsetki nieprawidłowych predykcji, w zależności od poziomów pomiaru zmiennych zależnych. Jeśli którakolwiek zmienna zależna ma poziom ilościowy, to wyświetlany jest średni błąd względny ogółem (względem średniej modelu). Jeśli wszystkie zmienne zależne są jakościowe, wyświetlany jest średni odsetek nieprawidłowych predykcji. Błędy względne lub odsetki nieprawidłowych predykcji są także prezentowane w odniesieniu do poszczególnych zmiennych zależnych.

- **Wyniki klasyfikacji.** Zawiera tabelę klasyfikacji dla każdej jakościowej zmiennej zależnej z podziałem na podzbiory i ogólnej. W każdej tabeli przedstawiona jest liczba obserwacji sklasyfikowanych prawidłowo i nieprawidłowo dla każdej kategorii zmiennej zależnej. Podawany jest także prawidłowo sklasyfikowany odsetek wszystkich obserwacji.
- **Krzywa ROC.** Przedstawia krzywą ROC (charakterystyka operacyjna odbiornika) dla każdej jakościowej zmiennej zależnej. Zawiera także tabelę z polami pod poszczególnymi krzywymi. Dla danej zmiennej zależnej wykres ROC zawiera po jednej krzywej dla każdej kategorii. Jeśli zmienna zależna ma dwie kategorie, to na każdej krzywej jej kategoria jest traktowana jako stan dodatni względem drugiej kategorii. Jeśli zmienna zależna ma więcej niż dwie kategorie, to na każdej krzywej jej kategoria jest traktowana jako stan dodatni względem agregacji wszystkich pozostałych kategorii.
- **Wykres skumulowanej korzyści.** Zawiera wykres skumulowanych korzyści dla każdej jakościowej zmiennej zależnej. Sposób prezentacji krzywych poszczególnych kategorii zmiennych zależnych jest taki sam, jak w przypadku krzywych ROC.
- **Wykres przyrostu.** Zawiera wykres przyrostu każdej jakościowej zmiennej zależnej. Sposób prezentacji krzywych poszczególnych kategorii zmiennych zależnych jest taki sam, jak w przypadku krzywych ROC.
- **Wykres przewidywanych przez obserwowane.** Zawiera wykres wartości przewidywanych wobec obserwowanych dla każdej zmiennej zależnej. W przypadku jakościowych zmiennych zależnych dla poszczególnych kategorii odpowiedzi wyświetlane są grupowane wykresy skrzynkowej przewidywanych pseudoprawdopodobieństw, z obserwowaną kategorią odpowiedzi jako zmienną grupującą. W przypadku ilościowych zmiennych zależnych wyświetlany jest wykres rozrzutu.
- **Wykres reszt przez wartości przewidywane.** Zawiera wykres reszt wobec wartości przewidywanych dla każdej ilościowej zmiennej zależnej. Między resztami a wartościami przewidywanymi nie powinno być widocznych zależności. Ten wykres jest generowany tylko dla ilościowych zmiennych zależnych.

Podsumowanie przetwarzania przypadku. Wyświetla tabelę podsumowania przetwarzania obserwacji, która podsumowuje liczbę obserwacji włączonych i wyłączonych z analizy, w sumie i według próby uczącej, testującej i wstrzymanej.

Analiza ważności zmiennych niezależnych. Przeprowadza analizę czułości, która oblicza znaczenie każdego predyktora dla wyznaczania sieci neuronowej. Analiza jest oparta na połączonej próbie uczącej i testującej lub tylko na próbie uczącej, jeśli nie ma próby testującej. W rezultacie powstaje tabela i wykres ważności oraz znormalizowanej ważności każdego predyktora. Należy zauważyć, że analiza czułości jest obliczeniowo kosztowna i czasochłonna, jeśli liczba predyktorów lub obserwacji jest duża.

Zapisywanie

Karta Zapisz służy do zapisywania predykcji jako zmiennych w zbiorze danych.

- **Zapisz przewidywaną wartość lub kategorię dla każdej zmiennej zależnej.** Zapisuje przewidywaną wartość ilościowych zmiennych zależnych i kategorii przewidywanych w przypadku jakościowych zmiennych zależnych.
- **Zapisz przewidywane pseudoprawdopodobieństwo lub kategorię dla każdej zmiennej zależnej.** Zapisuje pseudoprawdopodobieństwa dla jakościowych zmiennych zależnych. Dla każdej z pierwszych n kategorii zapisywana jest odrębna zmienna, gdzie n jest liczbą określoną w kolumnie **Kategorie do zapisania**.

Nazwy zapisywanych zmiennych. Automatyczne generowanie nazw zapewnia zachowanie wszystkich wyników pracy. Nazwy użytkownika umożliwiają likwidację/zastąpienie wyników poprzednich uruchomień bez wcześniejszego usuwania zapisanych zmiennych w Edytorze danych.

Prawdopodobieństwa i pseudoprawdopodobieństwa

Jakościowe zmienne zależne z aktywacją softmax i błędem entropii krzyżowej będą miały dla każdej kategorii wartość przewidywaną będącą prawdopodobieństwem tego, że dana obserwacja należy do tej kategorii.

Jakościowe zmienne zależne z błędem sumy kwadratów będą miały dla każdej kategorii wartość przewidywaną, której nie da się zinterpretować jako prawdopodobieństwa. Procedura zapisuje te pseudoprawdopodobieństwa nawet wtedy, gdy są mniejsze od 0 lub większe od 1, albo gdy ich suma dla danej zmiennej zależnej nie jest równa 1.

Na podstawie pseudoprawdopodobieństw tworzone są wykresy ROC, skumulowanych korzyści i wzrostu (patrz „Dane wyjściowe” na stronie 8). W przypadku, gdy dowolne z pseudoprawdopodobieństw jest mniejsze od 0 lub większe od 1 albo suma dla danej zmiennej nie jest równa 1, są one najpierw przeskalowywane, aby mieściły się w przedziale od 0 do 1 i sumowały się do wartości 1. Pseudoprawdopodobieństwa są przeskalowywane przez podzielenie przez ich sumę. Na przykład, jeśli obserwacja ma przewidywane pseudoprawdopodobieństwa 0,50, 0,60 i 0,40 dla zmiennej zależnej o trzech kategoriach, to każde pseudoprawdopodobieństwo jest dzielone przez sumę 1,50 w celu uzyskania wartości 0,33, 0,40 i 0,27.

Jeśli którekolwiek z pseudoprawdopodobieństw jest ujemne, to wartość bezwzględna najniższego z nich jest dodawana do wszystkich pseudoprawdopodobieństw przed powyższym przeskalowaniem. Na przykład, jeśli pseudoprawdopodobieństwa wynoszą -0,30, 0,50 i 1,30, to najpierw dodaje się 0,30 do każdej wartości w celu uzyskania 0,00, 0,80 i 1,60. Następnie każda nowa wartość jest dzielona przez sumę 2,40, co daje w wyniku 0,00, 0,33 i 0,67.

Eksport

Karta Eksport służy do zapisywania oszacowań wag synaptycznych poszczególnych zmiennych zależnych do pliku w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. . Ta opcja nie jest dostępna, jeśli zdefiniowano podzielone dane.

Opcje

Braki danych zdefiniowane przez użytkownika. Czynniki muszą posiadać prawidłowe wartości dla obserwacji, która ma zostać zawarta w analizie. Te elementy pozwalają zdecydować, czy wartości braków danych zdefiniowanych przez użytkownika są traktowane jako ważne wśród czynników i zależnych zmiennych jakościowych.

Reguły zatrzymujące. Są to reguły decydujące o momencie zakończenia uczenia sieci neuronowej. Uczenie obejmuje co najmniej jedno przejście przez dane. Następnie uczenie może zostać zakończone zgodnie z następującymi kryteriami, które sprawdzane są w kolejności, w jakiej zostały tutaj wymienione. W poniższych definicjach reguł zatrzymujących krok odpowiada przejściu przez dane w metodzie uczenia jednostkowo i podzbiorem oraz iteracji w metodzie uczenia całym zbiorem.

- **Maksymalna liczba kroków bez zmniejszenia błędu.** Liczba kroków dozwolona przed sprawdzeniem, czy błąd się zmniejszył. Jeśli po podanej liczbie kroków błąd się nie zmniejszy, uczenie zostaje zakończone. Podaj liczbę całkowitą większą od 0. Można również określić, która próbka danych jest używana do obliczenia błędu. **Wybierz automatycznie** powoduje użycie próby testującej, jeśli taka próba istnieje, a próby uczącej w przeciwnym razie. Należy zauważyć, że uczenie całym zbiorem gwarantuje zmniejszenie błędu dla próby uczącej po każdym przejściu przez dane; dlatego ta opcja ma zastosowanie do uczenia całym zbiorem tylko wtedy, gdy istnieje próba testująca. **Zarówno dane uczące i jak testowe** powoduje, że błąd jest sprawdzany dla obu prób; ta opcja ma zastosowanie tylko wtedy, gdy istnieje próba testująca.

Uwaga: Po każdym pełnym przejściu przez dane uczenie jednostkowo i podzbiorem wymaga dodatkowego przejścia przez dane w celu obliczenia błędu uczenia. To dodatkowe przejście przez dane może znacząco spowolnić uczenie, dlatego zasadniczo zaleca się zawsze udostępnienie próby testującej i wybranie opcji **Wybierz automatycznie**.

- **Maksymalny czas uczenia.** Ta opcja umożliwi określenie maksymalnej liczby minut, przez jaką będzie działał algorytm. Podaj liczbę większą od 0.
- **Maksymalna liczba epok uczących.** Dozwolona maksymalna liczba epok (przejdź przez dane). Przekroczenie maksymalnej liczby epok powoduje zakończenie uczenia. Podaj liczbę całkowitą większą od 0.
- **Minimalna względna zmiana błędu uczenia.** Uczenie kończy się, gdy względna zmiana błędu uczenia względem poprzedniego kroku jest mniejsza od wartości kryterium. Podaj liczbę większą od 0. W przypadku szkoleń on-line i jednostkowych kryterium to jest ignorowane, jeśli do obliczenia tego błędu używane są tylko dane testowe.
- **Minimalna względna zmiana współczynnika błędu uczenia.** Uczenie kończy się, gdy stosunek błędu uczenia do błędu modelu zerowego jest mniejszy od wartości kryterium. Model zerowy przewiduje wartość średnią wszystkich zmiennych zależnych. Podaj liczbę większą od 0. W przypadku szkoleń on-line i jednostkowych kryterium to jest ignorowane, jeśli do obliczenia tego błędu używane są tylko dane testowe.

Maksymalna liczba obserwacji zapisywana w pamięci. Steruje następującymi ustawieniami algorytmów perceptronu wielowarstwowego. Podaj liczbę całkowitą większą od 1.

- Przy automatycznym wyborze architektury wielkość próby używanej do wyboru architektury wynosi $\min(1000, memsize)$, gdzie *memsize* jest maksymalną liczbą obserwacji, jaka może być przechowywana w pamięci.
- W przypadku uczenia podzbiorem z automatycznym obliczaniem liczby podzbiorów ta liczba podzbiorów wynosi $\min(\max(M/10, 2), memsize)$, gdzie *M* jest liczbą obserwacji w próbie uczącej.

radialna funkcja bazowa

Radialna funkcja bazowa (RBF) umożliwia utworzenie modelu predykcyjnego do co najmniej jednej zmiennej zależnej (wynikowej) w oparciu o wartości zmiennych predykcyjnych.

Przykład. Operator telekomunikacyjny pogrupował bazę klientów wg wzorców korzystania z usług, tworząc cztery kategorie. Sieć RBF korzystająca z danych demograficznych do przewidywania członkostwa w grupach umożliwia firmie dostosowanie ofert do potencjalnych przyszłych klientów indywidualnych.

Zagadnienia dotyczące danych

Zmienne zależne. Możliwe zmienne zależne to:












- *Nominalny.* Zmienna może być traktowana jako nominalna, gdy jej wartości reprezentują kategorie bez wewnętrznego rankingu (na przykład dział przedsiębiorstwa, w którym pracuje pracownik). Przykładami zmiennych nominalnych są: region, kod pocztowy lub wyznanie.
- *Porządkowy.* Zmienna może być traktowana jako porządkowa, gdy jej wartości reprezentują kategorie z jakimś nieodłącznym rangą (na przykład poziomy zadowolenia z usługi z bardzo niezadowolonego do bardzo zadowolonego). Przykładami zmiennych porządkowych mogą być oceny opinii reprezentujące stopień satysfakcji lub przekonania oraz oceny preferencji.

- *scale*. Zmienna może być traktowana jako skala (ilościowa), gdy jej wartości reprezentują uporządkowane kategorie z miarodajnym pomiarem, tak aby porównania odległości między wartościami były odpowiednie. Przykładami zmiennych ilościowych mogą być: wiek w latach lub przychód w tysiącach złotych.

W procedurze przyjmuje się, że odpowiedni poziom pomiaru został przypisany do wszystkich zmiennych zależnych, choć można tymczasowo zmienić poziom pomiaru dla zmiennej, klikając prawym przyciskiem myszy zmienną na liście zmiennych źródłowych i wybierając poziom pomiaru z menu kontekstowego.

Ikona obok każdej zmiennej na liście zmiennych określa poziom pomiaru oraz typ danych:

Tabela 2. Ikony poziomu pomiaru

	Liczbowy	Łańcuch	Data	Czas
Zmienna (ilościowa)		nie dotyczy		
Porządkowy				
Nominalna				

Zmienne predykcyjne. Predyktory można określić jako czynniki (jakościowe) lub współzmienne (skala).

Kodowanie zmiennych jakościowych. Procedura tymczasowo przekodowuje predyktory jakościowe i zmienne zależne, używając jednego z c kodowań na czas trwania procedury. Jeśli istnieje c kategorii zmiennej, to zmienna jest przechowywana jako c wektorów, z pierwszą kategorią oznaczoną $(1, 0, \dots, 0)$, następną kategorią $(0, 1, 0, \dots, 0)$, ..., i kategorii końcowej $(0, 0, \dots, 0, 1)$.

Ten schemat kodowania zwiększa liczbę wag synaptycznych i może spowodować spowolnienie uczenia; jednak zastosowanie mniej złożonych metod kodowania zwykle prowadzi do słabszego dopasowania sieci neuronowych. Jeśli uczenie sieci przebiega bardzo wolno, spróbuj zmniejszyć liczbę kategorii w predyktorach kategoryalnych, łącząc podobne kategorie lub usuwając przypadki, które mają bardzo rzadkie kategorie.

Kodowanie jeden z c jest oparte na danych szkoleniowych, nawet jeśli zdefiniowano próbę testowania lub wstrzymania (patrz "Podzbiory" na stronie 13). Dlatego jeśli próba testowania lub wstrzymania zawiera obserwacje z kategoriami predyktorów, które nie są obecne w danych uczących, wówczas obserwacje te nie są używane przez procedurę ani oceniane. Jeśli próba testowania lub wstrzymania zawiera obserwacje z kategoriami zmiennych zależnych, które nie są obecne w danych uczących, wówczas obserwacje te nie są używane przez procedurę, ale mogą być oceniane.

Przeskalowanie. Ilościowe zmienne zależne i współzmienne zostają domyślnie przeskalowane, aby usprawnić uczenie sieci. Przeskalowanie odbywa się na podstawie danych uczących, nawet jeśli zdefiniowana została próba testowania lub wstrzymania (patrz "Podzbiory" na stronie 13). Oznacza to, że w zależności od typu przeskalowania średnia, odchylenie standardowe, wartość minimalna lub wartość maksymalna współzmiennnej i zmiennej zależnej są obliczane z użyciem wyłączenie danych uczących. Jeżeli zostanie określona zmienna definiująca podziały, ważne jest, aby współzmienne i zmienne zależne miały podobny rozkład w próbach uczących, testowania i wstrzymania.

Wagi liczebności. Wagi częstotliwości są ignorowane przez tę procedurę.

Replikacja wyników. Aby przeprowadzić dokładną replikację wyników, należy użyć tej samej wartości inicjalizacji dla generatora liczb losowych, tej samej kolejności danych oraz użyć tych samych ustawień procedury. Aby uzyskać więcej szczegółów na temat tego problemu, należy zapoznać się z tematami:

- **Generowanie liczb losowych.** Procedura używa generatora liczb losowych przy losowym przydziale podziałów. Aby ponownie utworzyć takie same losowe wyniki w przyszłości, należy użyć tej samej wartości inicjalizacji dla generatora liczb losowych przed każdym uruchomieniem procedury radialnej funkcji bazowej.

- **Kolejność obserwacji.** Wyniki są również zależne od kolejności danych, ponieważ algorytm dwustopniowej analizy skupień służy do określenia bazowych funkcji radialnych.

Aby zminimalizować wpływ kolejności, należy losowo ustawić obserwacje. Aby sprawdzić stabilność danego rozwiązania, można uzyskać kilka różnych rozwiązań z obserwacjami posortowanymi w różnej, losowej kolejności. W sytuacji, gdy wielkość pliku jest ekstremalnie duża, wykonywanych jest wiele uruchomień z użyciem próby obserwacji posortowanych w różnej kolejności losowej.

Tworzenie sieci radialnej funkcji bazowej

Z menu wybierz:

Analiza > Sieci neuronowe > Radialna funkcja bazowa...

1. Wybierz co najmniej jedną zmienną zależną.
2. Wybierz co najmniej jeden czynnik lub współzmienną.

Opcjonalnie, na karcie Zmienne można zmienić metodę przeskalowania współzmiennych. Dostępne są następujące opcje:

- **Standaryzowane.** Odjęcie średniej i podział przez standardowe odchylenie, $(x - \text{średnia})/s$.
- **Znormalizowane.** Odjęcie wartości minimalnej i podział przez zakres, $(x - \text{min.})/(\text{maks.} - \text{min.})$. Znormalizowane wartości należą do przedziału od 0 do 1.
- **Skorygowane znormalizowane.** Skorygowana wersja odejmowania wartości minimalnej i podziału przez zakres, $[2 * (x - \text{min.})/(\text{maks.} - \text{min.})] - 1$. Skorygowane wartości znormalizowane zawierają się w zakresie od -1 do 1.
- **Brak.** Bez przeskalowania współzmiennych.

Zmienne z nieznanym poziomem pomiaru

Alert poziomu pomiaru wyświetla się, gdy poziom pomiaru dla jednej lub większej ilości zmiennych w zbiorze danych jest nieznan. Ponieważ poziom pomiaru wpływa na wyliczenie wyników dla tej procedury, wszystkie zmienne muszą mieć zdefiniowany poziom pomiaru.

Skanowanie danych. Odczytuje dane w aktywnym zbiorze danych i przypisuje domyślny poziom pomiaru do wszystkich zmiennych, które mają aktualnie nieznan poziom pomiaru. Jeśli zbiór danych jest duży, może to zająć trochę czasu.

Przypisz ręcznie. Otwiera okno dialogowe, które zestawia wszystkie zmienne z nieznanym poziomem pomiaru. Można użyć tego okna dialogowego do przypisania poziomu pomiaru do tych zmiennych. Można również przypisać poziom pomiaru w Widoku zmiennych Edytora danych.

Ponieważ poziom pomiaru jest ważny dla tej procedury, nie można wejść do tego okna dialogowego w celu uruchomienia tej procedury, dopóki wszystkie zmienne nie będą miały zdefiniowanego poziomu pomiaru.

Podzbiory

Podzbiór danych. Ta grupa określa metodę podziału aktywnego zbioru danych na próby uczące, testu i wstrzymane. **Próba ucząca** składa się z rekordów danych używanych do szkolenia sieci neuronowych. Pewna wartość procentowa obserwacji w zbiorze danych musi być przypisana do próby uczącej w celu uzyskania modelu. **Próba testu** to niezależny zbiór danych używanych do nadzorowania błędów w czasie uczenia. Pozwala to zapobiec nadmiernemu wyuczeniu. Zdecydowanie zaleca się utworzenie próby testowej, a szkolenie sieci będzie bardziej wydajne, jeśli próba testu będzie mniejsza niż próba testująca. **Próba wstrzymana** jest innym niezależnym zestawem rekordów danych używanych do oceny końcowej sieci neuronowych. Błąd w próbie wstrzymanej daje „uczciwą” ocenę możliwości predykcyjnej modelu, ponieważ obserwacje wstrzymane nie zostały użyte do budowy modelu.

- **Przypisz obserwacje losowo na bazie względnej liczby obserwacji.** Należy określić względną liczbę (stosunek) obserwacji losowo przypisanych do każdej próby (uczącej, testu i wstrzymanej). W kolumnie % wskazany jest odsetek obserwacji, które zostaną przypisane do każdej próby w oparciu o podane liczby względne.

Na przykład, podanie 7, 3, 0 jako względnych liczb dla obserwacji w próbie uczącej, testu i wstrzymanej odpowiada odsetkom 70%, 30% i 0%. Podanie 2, 1, 1 jako względnych liczb odpowiada 50%, 25% i 25%; 1, 1, 1 powoduje podział zbioru danych na trzy równe części między próbę uczącą, testu i wstrzymaną.

- **Użyj zmiennej dzielącej w celu przydzielenia obserwacji.** Określ zmienną numeryczną, która przydziela każdą obserwację w aktywnym zbiorze danych do próby uczącej, testu lub wstrzymanej. Obserwacje z wartością pozytywną są przydzielane do próby uczącej, a przypadki z wartością 0 lub mniejszą — do próby wstrzymanej. Obserwacje z systemowymi brakami danych zostają wykluczone z analizy. Wartości zdefiniowanych przez użytkownika braków danych zmiennej dzielącej zawsze są traktowane jako prawidłowe.

Architektura

Karta Architektura służy do określania struktury sieci. Procedura tworzy sieć neuronową z jedną ukrytą warstwą „radialnej funkcji bazowej”; z reguły nie będzie potrzeby zmiany tych ustawień.

Liczba neuronów w warstwie ukrytej. Istnieją trzy sposoby wybierania liczby ukrytych jednostek.

1. **Znajdź najlepszą liczbę neuronów w automatycznie obliczonym zakresie wartości.** Procedura automatycznie oblicza minimalną i maksymalną wartość w zakresie i znajduje najlepszą liczbę ukrytych jednostek w zakresie.

Jeśli zdefiniowana jest próba testująca, procedura używa kryterium danych testujących: najlepsza liczba ukrytych jednostek to taka, która daje najmniejszy błąd w danych testujących. Jeśli próba testująca nie jest zdefiniowana, procedura używa kryterium informacyjnego Bayesa (BIC): najlepsza liczba ukrytych jednostek to taka, która daje najmniejszą wartość BIC na podstawie danych uczących.
2. **Znajdź najlepszą liczbę neuronów w zakresie wartości.** Można podać własny zakres, a procedura znajdzie „najlepszą” liczbę ukrytych jednostek w tym zakresie. Tak jak poprzednio, najlepsza liczba jednostek z zakresu wyznaczana jest na podstawie kryterium danych testujących albo kryterium BIC.
3. **Użyj określonej liczby neuronów.** Można bezpośrednio podać konkretną liczbę jednostek i nie używać zakresu.

Funkcja aktywacji dla warstwy ukrytej. Funkcja aktywacji dla ukrytej warstwy jest radialną funkcją bazową, która służy do łączenia jednostek w warstwie do wartości jednostek w następnej warstwie. Dla warstwy wyjściowej funkcja aktywacji jest funkcją tożsamości. Jednostki wyjściowe to sumy ważone ukrytych jednostek.

- **Znormalizowana radialna funkcja bazowa.** Powoduje użycie funkcji aktywacji softmax, tak by aktywacje wszystkich jednostek ukrytych były znormalizowane i sumowały się do 1.
- **Zwykła radialna funkcja bazowa.** Używa wykładniczej funkcji aktywacji, tak by aktywacja ukrytej jednostki była „garbem” rozkładu Gaussa w funkcji danych wejściowych..

Pokrycie w warstwie ukrytej. Współczynnik pokrycia jest mnożnikiem stosowanym do szerokości bazowych funkcji radialnych. Automatycznie obliczona wartość współczynnika pokrycia wynosi $1+0,1d$, gdzie d jest liczbą jednostek wejściowych (sumą liczby kategorii wszystkich czynników i liczby współzmiennych).

Dane wyjściowe

Struktura sieci. Wyświetla informacje podsumowujące na temat sieci neuronowej.

- **Opis.** Wyświetla informacje na temat sieci neuronowej, w tym zmienne zależne, liczbę jednostek wejściowych i wyjściowych, liczbę ukrytych warstw i jednostek oraz funkcje aktywacji.
- **Diagram.** Wyświetla diagram sieci w postaci nieedytowalnej. Należy zwrócić uwagę, że w miarę wzrostu liczby współzmiennych i poziomów czynników diagram staje się coraz trudniejszy do zinterpretowania.
- **Wagi synaptyczne.** Wyświetla oszacowania współczynników, które odzwierciedlają relację między jednostkami na danej warstwie a jednostkami na następnej warstwie. Wagi synaptyczne są oparte na próbie uczącej, nawet jeśli aktywny zbiór danych jest podzielony na dane uczące, testujące

i wstrzymane. Należy zwrócić uwagę, że liczba wag synaptycznych może być duża, a te wagi co do zasady nie służą do interpretacji wyników działania sieci.

Wydażność sieci. Przedstawia wyniki służące do określenia, czy model jest „dobry”. *Uwaga:* Wykresy w tej grupie są oparte na połączonej próbie uczącej i testującej lub tylko na próbie uczącej, jeśli nie ma próby testującej.

- **Podsumowanie modelu.** Przedstawia wyniki działania sieci neuronowej z podziałem na grupy danych oraz wyniki łączne, w tym błąd, błąd względny lub odsetek nieprawidłowych predykcji oraz czas uczenia.

Jest to błąd sumy kwadratów. Ponadto prezentowane są błędy względne lub odsetki nieprawidłowych predykcji, w zależności od poziomów pomiaru zmiennych zależnych. Jeśli którakolwiek zmienna zależna ma poziom ilościowy, to wyświetlany jest średni błąd względny ogółem (względem średniej modelu). Jeśli wszystkie zmienne zależne są jakościowe, wyświetlany jest średni odsetek nieprawidłowych predykcji. Błędy względne lub odsetki nieprawidłowych predykcji są także prezentowane w odniesieniu do poszczególnych zmiennych zależnych.

- **Wyniki klasyfikacji.** Zawiera tabelę klasyfikacji dla każdej jakościowej zmiennej zależnej. W każdej tabeli przedstawiona jest liczba obserwacji sklasyfikowanych prawidłowo i nieprawidłowo dla każdej kategorii zmiennej zależnej. Podawany jest także prawidłowo sklasyfikowany odsetek wszystkich obserwacji.
- **Krzywa ROC.** Przedstawia krzywą ROC (charakterystyka operacyjna odbiornika) dla każdej jakościowej zmiennej zależnej. Zawiera także tabelę z polami pod poszczególnymi krzywymi. Dla danej zmiennej zależnej wykres ROC zawiera po jednej krzywej dla każdej kategorii. Jeśli zmienna zależna ma dwie kategorie, to na każdej krzywej jej kategoria jest traktowana jako stan dodatni względem drugiej kategorii. Jeśli zmienna zależna ma więcej niż dwie kategorie, to na każdej krzywej jej kategoria jest traktowana jako stan dodatni względem agregacji wszystkich pozostałych kategorii.
- **Wykres skumulowanej korzyści.** Zawiera wykres skumulowanych korzyści dla każdej jakościowej zmiennej zależnej. Sposób prezentacji krzywych poszczególnych kategorii zmiennych zależnych jest taki sam, jak w przypadku krzywych ROC.
- **Wykres przyrostu.** Zawiera wykres przyrostu każdej jakościowej zmiennej zależnej. Sposób prezentacji krzywych poszczególnych kategorii zmiennych zależnych jest taki sam, jak w przypadku krzywych ROC.
- **Wykres przewidywanych przez obserwowane.** Zawiera wykres wartości przewidywanych wobec obserwowanych dla każdej zmiennej zależnej. W przypadku jakościowych zmiennych zależnych dla poszczególnych kategorii odpowiedzi wyświetlane są grupowane wykresy skrzynkowej przewidywanych pseudoprawdopodobieństw, z obserwowaną kategorią odpowiedzi jako zmienną grupującą. W przypadku ilościowych zmiennych zależnych wyświetlany jest wykres rozrzutu.
- **Wykres reszt przez wartości przewidywane.** Zawiera wykres reszt wobec wartości przewidywanych dla każdej ilościowej zmiennej zależnej. Między resztami a wartościami przewidywanymi nie powinno być widocznych zależności. Ten wykres jest generowany tylko dla ilościowych zmiennych zależnych.

Podsumowanie przetwarzania przypadku. Wyświetla tabelę podsumowania przetwarzania obserwacji, która podsumowuje liczbę obserwacji włączonych i wyłączonej z analizy, w sumie i według próby uczącej, testującej i wstrzymanej.

Analiza ważności zmiennych niezależnych. Przeprowadza analizę czułości, która oblicza znaczenie każdego predyktora dla wyznaczania sieci neuronowej. Analiza jest oparta na połączonej próbie uczącej i testującej lub tylko na próbie uczącej, jeśli nie ma próby testującej. W rezultacie powstaje tabela i wykres ważności oraz znormalizowanej ważności każdego predyktora. Należy zauważyć, że analiza czułości jest obliczeniowo kosztowna i czasochłonna, jeśli liczba predyktorów lub obserwacji jest duża.

Zapisywanie

Karta Zapisz służy do zapisywania predykcji jako zmiennych w zbiorze danych.

- **Zapisz przewidywaną wartość lub kategorię dla każdej zmiennej zależnej.** Zapisuje przewidywaną wartość ilościowych zmiennych zależnych i kategorii przewidywanych w przypadku jakościowych zmiennych zależnych.

- **Zapisz przewidywane pseudoprawdopodobieństwo dla każdej zmiennej zależnej.** Zapisuje pseudoprawdopodobieństwa dla jakościowych zmiennych zależnych. Dla każdej z pierwszych n kategorii zapisywana jest odrębna zmienna, gdzie n jest liczbą określoną w kolumnie *Kategorie do zapisania*.

Nazwy zapisywanych zmiennych. Automatyczne generowanie nazw zapewnia zachowanie wszystkich wyników pracy. Nazwy użytkownika umożliwiają likwidację lub zastąpienie wyników poprzednich uruchomień bez wcześniejszego usuwania zapisanych zmiennych w Edytorze danych.

Prawdopodobieństwa i pseudoprawdopodobieństwa

Przewidywane pseudoprawdopodobieństwa nie mogą być interpretowane jako prawdopodobieństwa, ponieważ procedura Radialna funkcja bazowa używa błędu sumy kwadratów i tożsamosciowej funkcji aktywacji dla warstwy wyjściowej. Procedura zapisuje te pseudoprawdopodobieństwa nawet wtedy, gdy są mniejsze od 0 lub większe od 1, albo gdy ich suma dla danej zmiennej zależnej nie jest równa 1.

Na podstawie pseudoprawdopodobieństw tworzone są wykresy ROC, skumulowanych korzyści i wzrostu (patrz ["Dane wyjściowe"](#) na stronie 14). W przypadku, gdy dowolne z pseudoprawdopodobieństw jest mniejsze od 0 lub większe od 1 albo suma dla danej zmiennej nie jest równa 1, są one najpierw przeskalowywane, aby mieściły się w przedziale od 0 do 1 i sumowały się do wartości 1. Pseudoprawdopodobieństwa są przeskalowywane przez podzielenie przez ich sumę. Na przykład, jeśli obserwacja ma przewidywane pseudoprawdopodobieństwa 0,50, 0,60 i 0,40 dla zmiennej zależnej o trzech kategoriach, to każde pseudoprawdopodobieństwo jest dzielone przez sumę 1,50 w celu uzyskania wartości 0,33, 0,40 i 0,27.

Jeśli którekolwiek z pseudoprawdopodobieństw jest ujemne, to wartość bezwzględna najniższego z nich jest dodawana do wszystkich pseudoprawdopodobieństw przed powyższym przeskalowaniem. Na przykład, jeśli pseudoprawdopodobieństwa wynoszą -0,30, .50 i 1,30, to najpierw dodaje się 0,30 do każdej wartości w celu uzyskania 0,00, 0,80 i 1,60. Następnie każda nowa wartość jest dzielona przez sumę 2,40, co daje w wyniku 0,00, 0,33 i 0,67.

Eksport

Karta Eksport służy do zapisywania oszacowań wag synaptycznych poszczególnych zmiennych zależnych do pliku w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. . Ta opcja nie jest dostępna, jeśli zdefiniowano podzielone dane.

Opcje

Braki danych zdefiniowane przez użytkownika. Czynniki muszą posiadać prawidłowe wartości dla obserwacji, która ma zostać zawarta w analizie. Te elementy pozwalają zdecydować, czy wartości braków danych zdefiniowanych przez użytkownika są traktowane jako ważne wśród czynników i zależnych zmiennych jakościowych.

Uwagi

Niniejsza publikacja została przygotowana z myślą o produktach i usługach oferowanych w Stanach Zjednoczonych. IBM może udostępniać ten materiał w innych językach. Jednakże w celu uzyskania dostępu do takiego materiału istnieje konieczność posiadania egzemplarza produktu w takim języku.

Produktów, usług lub opcji opisywanych w tym dokumencie IBM nie musi oferować we wszystkich krajach. Informacje o produktach i usługach dostępnych w danym kraju można uzyskać od lokalnego przedstawiciela IBM. Odwołanie do produktu, programu lub usługi IBM nie oznacza, że można użyć wyłącznie tego produktu, programu lub usługi IBM. Zamiast nich można zastosować ich odpowiednik funkcjonalny pod warunkiem że nie narusza to praw własności intelektualnej IBM. Jednakże cała odpowiedzialność za ocenę przydatności i sprawdzenie działania produktu, programu lub usługi pochodzących od producenta innego niż IBM spoczywa na użytkowniku.

IBM może posiadać patenty lub złożone wnioski patentowe na produkty, o których mowa w niniejszej publikacji. Przedstawienie tej publikacji nie daje żadnych uprawnień licencyjnych do tychże patentów. Pisemne zapytania w sprawie licencji można przesyłać na adres:

IBM Director of Licensing

IBM Corporation

*North Castle Drive, MD-NC119
Armonk, NY 10504-1785 U.S.A.*

Zapytania dotyczące zestawów znaków dwubajtowych (DBCS) należy kierować do lokalnych działów własności intelektualnej IBM (IBM Intellectual Property Department) lub wysłać je na piśmie na adres:

Intellectual Property Licensing

*Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokio 103-8510, Japonia*

INTERNATIONAL BUSINESS MACHINES CORPORATION DOSTARCZA TĘ PUBLIKACJĘ W STANIE, W JAKIM SIĘ ZNAJDUJE ("AS IS") BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI (W TYM TAKŻE RĘKOJMI), WYRAŻNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI, ŻE PUBLIKACJA NIE NARUSZA PRAW STRON TRZECICH. Ustawodawstwa niektórych krajów nie dopuszczają zastrzeżeń dotyczących gwarancji wyraźnych lub domniemanych w odniesieniu do pewnych transakcji; w takiej sytuacji powyższe zdanie nie ma zastosowania.

Informacje zawarte w tej publikacji mogą zawierać nieścisłości techniczne lub błędy drukarskie. Informacje te są okresowo aktualizowane, a zmiany te zostaną uwzględnione w kolejnych wydaniach tej publikacji. IBM zastrzega sobie prawo do wprowadzania ulepszeń i/lub zmian w produktach i/lub programach opisanych w tej publikacji w dowolnym czasie, bez wcześniejszego powiadomienia.

Wszelkie wzmianki w tej publikacji na temat stron internetowych firm innych niż IBM zostały wprowadzone wyłącznie dla wygody użytkowników i w żadnym razie nie stanowią zachęty do ich odwiedzania. Materiały dostępne na tych stronach nie są częścią materiałów opracowanych dla tego produktu IBM, a użytkownik korzysta z nich na własną odpowiedzialność.

IBM ma prawo do używania i rozpowszechniania informacji przystanych przez użytkownika w dowolny sposób, jaki uzna za właściwy, bez żadnych zobowiązań wobec ich autora.

Licencjodawcy tego programu, którzy chcieliby uzyskać informacje na temat programu w celu: (i) umożliwienia wymiany informacji między niezależnie utworzonymi programami i innymi programami

(łącznie z opisywanym) oraz (ii) wykorzystywania wymienianych informacji, powinni skontaktować się z:

IBM Director of Licensing

IBM Corporation

*North Castle Drive, MD-NC119
Armonk, NY 10504-1785 U.S.A.*

Informacje takie mogą być udostępnione, o ile spełnione zostaną odpowiednie warunki, w tym, w niektórych przypadkach, zostanie uiszczona stosowna opłata.

Licencjonowany program opisany w niniejszej publikacji oraz wszystkie inne licencjonowane materiały dostępne dla tego programu są dostarczane przez IBM na warunkach określonych w Umowie IBM z Klientem, Międzynarodowej Umowie Licencyjnej IBM na Program lub w innych podobnych umowach zawartych między IBM i użytkownikami.

Dane dotyczące wydajności i cytowane przykłady zostały przedstawione jedynie w celu zobrazowania sytuacji. Faktyczne wyniki dotyczące wydajności mogą się różnić w zależności do konkretnych warunków konfiguracyjnych i operacyjnych.

Informacje dotyczące produktów innych podmiotów niż IBM zostały uzyskane od dostawców tych produktów, z ich publicznych ogłoszeń lub innych dostępnych publicznie źródeł. IBM nie testował tych produktów i nie może potwierdzić dokładności pomiarów wydajności, kompatybilności ani żadnych innych danych związanych z produktami firm innych niż IBM. Pytania dotyczące możliwości produktów firm innych niż IBM należy kierować do dostawców tych produktów.

Wszelkie stwierdzenia dotyczące przyszłych kierunków rozwoju i zamierzeń IBM mogą zostać zmienione lub wycofane bez powiadomienia.

Publikacja ta zawiera przykładowe dane i raporty używane w codziennych operacjach działalności gospodarczej. W celu kompleksowego zilustrowania tej działalności podane przykłady zawierają nazwy osób, firm i ich produktów. Wszystkie te nazwy/nazwiska są fikcyjne i jakiegokolwiek podobieństwo do istniejących nazw/nazwisk jest całkowicie przypadkowe.

LICENCJA W ZAKRESIE PRAW AUTORSKICH:

Niniejsza publikacja zawiera przykładowe aplikacje w kodzie źródłowym ilustrujące techniki programowania w różnych systemach operacyjnych. Użytkownik może kopiować, modyfikować i rozpowszechniać te programy przykładowe w dowolnej formie bez uiszczania opłat na rzecz IBM, w celu rozbudowy, użytkowania, handlowego lub w celu rozpowszechniania aplikacji zgodnych z aplikacyjnym interfejsem programowym dla tego systemu operacyjnego, dla którego napisane były programy przykładowe. Programy przykładowe nie zostały gruntownie przetestowane. IBM nie może zatem gwarantować ani sugerować niezawodności, użyteczności i funkcjonalności tych programów. Programy przykładowe są dostarczane w stanie, w jakim się znajdują ("AS IS"), bez jakichkolwiek gwarancji (rękojmię również wyłącza się). IBM nie ponosi odpowiedzialności za jakiegokolwiek szkody wynikające z używania programów przykładowych.

Każda kopia programu przykładowego lub jakiegokolwiek jego fragment, jak też jakiegokolwiek prace pochodne muszą zawierać następujące uwagi dotyczące praw autorskich:

© Copyright IBM Corp. 2021. Fragmenty tego kodu pochodzą z przykładowych programów produktu IBM Corp. Programy przykładowe.

© Copyright IBM Corp. 1989-2021. Wszelkie prawa zastrzeżone.

Znaki towarowe

IBM, logo IBM i ibm.com są znakami towarowymi lub zastrzeżonymi znakami towarowymi International Business Machines Corp., zarejestrowanymi w wielu systemach prawnych na całym świecie. Pozostałe nazwy produktów i usług mogą być znakami towarowymi IBM lub innych przedsiębiorstw. Aktualna lista znaków towarowych IBM dostępna jest w serwisie WWW, w sekcji "Copyright and trademark

information" (Informacje o prawach autorskich i znakach towarowych), pod adresem www.ibm.com/legal/copytrade.shtml.

Adobe, logo Adobe, PostScript oraz logo PostScript są znakami towarowymi lub zastrzeżonymi znakami towarowymi Adobe Systems Incorporated w Stanach Zjednoczonych i/lub w innych krajach.

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium i Pentium są znakami towarowymi lub zastrzeżonymi znakami towarowymi Intel Corporation lub przedsiębiorstw podporządkowanych w Stanach Zjednoczonych i w innych krajach.

Linux jest zastrzeżonym znakiem towarowym Linusa Torvaldsa w Stanach Zjednoczonych i/lub w innych krajach.

Microsoft, Windows, Windows NT oraz logo Windows są znakami towarowymi Microsoft Corporation w Stanach Zjednoczonych i/lub w innych krajach.

UNIX jest zastrzeżonym znakiem towarowym Open Group w Stanach Zjednoczonych i w innych krajach.

Java oraz wszystkie znaki towarowe i logo dotyczące Java są znakami towarowymi firmy i jej firm zależnych.

Indeks

A

architektura
 sieci neuronowe [2](#)
architektura sieci
 w perceptronie wielowarstwowym [6](#)
 w radialnej funkcji bazowej [14](#)

B

brakujące wartości
 w perceptronie wielowarstwowym [10](#)

D

diagram struktury sieci
 w perceptronie wielowarstwowym [8](#)
 w radialnej funkcji bazowej [14](#)

F

funkcja aktywacji
 w perceptronie wielowarstwowym [6](#)
 w radialnej funkcji bazowej [14](#)

K

Krzywa ROC
 w perceptronie wielowarstwowym [8](#)
 w radialnej funkcji bazowej [14](#)

P

Perceptron wielowarstwowo
 architektura sieci [6](#)
 eksport modelu [10](#)
 opcje [10](#)
 partycje [5](#)
 uczenie [7](#)
 wyniki [8](#)
 zapisywanie zmiennych w aktywnym zbiorze danych [10](#)
Perceptron wielowarstwowo (MLP) [3](#)
próba szkoleniowa
 w perceptronie wielowarstwowym [5](#)
 w radialnej funkcji bazowej [13](#)
próba testująca
 w perceptronie wielowarstwowym [5](#)
 w radialnej funkcji bazowej [13](#)
próba wstrzymana
 w perceptronie wielowarstwowym [5](#)
 w radialnej funkcji bazowej [13](#)

R

radialna funkcja bazowa
 opcje [16](#)

radialna funkcja bazowa (*kontynuacja*)
 partycje [13](#)
 wyniki [14](#)
 zapisywanie zmiennych w aktywnym zbiorze danych [15](#)
Radialna funkcja bazowa
 architektura sieci [14](#)
 eksport modelu [16](#)
reguły zatrzymujące
 w perceptronie wielowarstwowym [10](#)

S

sieci neuronowe
 architektura [2](#)

U

uczenie całym zbiorem
 w perceptronie wielowarstwowym [7](#)
uczenie jednostkowo
 w perceptronie wielowarstwowym [7](#)
uczenie podzbiorem
 w perceptronie wielowarstwowym [7](#)
uczenie sieci
 w perceptronie wielowarstwowym [7](#)

W

warstwa ukryta
 w perceptronie wielowarstwowym [6](#)
 w radialnej funkcji bazowej [14](#)
warstwa wyjściowa
 w perceptronie wielowarstwowym [6](#)
 w radialnej funkcji bazowej [14](#)
wykres korzyści
 w perceptronie wielowarstwowym [8](#)
 w radialnej funkcji bazowej [14](#)
wykres przyrostu
 w perceptronie wielowarstwowym [8](#)
 w radialnej funkcji bazowej [14](#)

