

Sztuczne sieci neuronowe - założenia projektu

Aleksandra Poręba Grzegorz Podsiadło

17 czerwca 2020

Wydział Fizyki i Informatyki Stosowanej
ul. Reymonta 19
30-055 Kraków
Polska

1. Wstęp
2. Zbiór danych
3. Problem
4. Planowane rozwiązania
5. Rozwiązania
6. Podsumowanie
7. Bibliografia

Wstęp

Tematem naszego projektu jest przewidzenie wyniku egzaminu SAT na podstawie czynników środowiskowych.

Wybrany zbiór danych pozwoli na przeprowadzenie kompleksowej analizy problemu z wykorzystaniem wielu poznanych technik związanych ze sztucznymi sieciami neuronowymi.

Zbiór danych

Zbiór danych, który zostanie użyty przy rozwiązywaniu problemu pochodzi z repozytorium *kaggle.com* [4], dostępnym pod [adresem](#).

Składa się on z 8 kolumn, określających:

- Płeć,
- Rasę,
- Wykształcenie rodzica,
- Przystąpienie do kursu powiązanego z testem,
- Rodzaj diety dostarczanej przez szkołę,
- Wynik egzaminu SAT z matematyki,
- Wynik egzaminu SAT z czytania,
- Wynik egzaminu SAT z pisanie.

Problem

Podczas pracy nad projektem będziemy szukać odpowiedzi na pytanie które czynniki mają największy wpływ na wynik testu.

Badania pozwolą nam określić, które czynniki możemy odrzucić przy przewidywaniu wyników dla danych egzaminów, a które mają istotny wpływ.

Zostanie również zbadane czy jesteśmy w stanie przewidzieć wynik z zadowalającą dokładnością tylko na podstawie znajomości rezultatów z dwóch pozostałych egzaminów.

Planowane rozwiązania

Projekt zrealizowany zostanie w oparciu o środowisko Matlab.

Planujemy zbadanie tematu pod różnymi kątami, z wykorzystaniem różnych rodzajów sieci do klasyfikacji, między innymi:

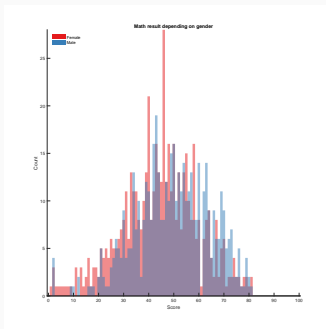
- Jednowarstwowe sieci neuronów dyskretnych
- Sieć perceptronów wielowarstwowych (ang. MLP)

Zostaną przetestowane różne parametry owych sieci, ilości neuronów oraz liczebności zbiorów uczących.

Rozwiązania

Analiza zbioru danych

Przed przystąpieniem do tworzenia sieci postanowiono na dokonanie analizy zbioru dostarczonych danych. W tym celu utworzono szereg różnych histogramów pozwalających na sprawdzenie zależności między kolumnami danych a wynikami egzaminu.



Rysunek 1: Przykład porównania wyników z matematyki w zależności od płci.

Utrzymane rozkłady wyników egzaminów były zbliżone do rozkładu normalnego. Przeprowadzone badanie pokazało, że mężczyźni uzyskują lepsze rezultaty w części matematycznej, kobiety natomiast w pozostałych dwóch częściach.

Dodatkowo widoczny jest silny pozytywny wpływ przystąpienia do kursu wstępnego, oraz mały wpływ przyjmowanego posiłku. Widoczny był również wpływ wykształcenia rodziców.

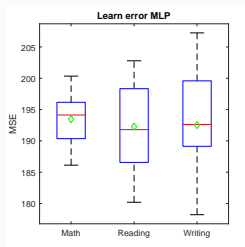
W dalszej części została zbadana korelacja pomiędzy danymi. Największa zależność została zaobserwowana pomiędzy wynikami egzaminów - można spodziewać się, że sieć będzie osiągać lepsze rezultaty przy uczeniu nimi.

Dla czynników środowiskowych zależności były niewielkie - poniżej wartości 0.5. Na podstawie współczynnika korelacji można spodziewać się, że znaczenie będzie mieć fakt uczęszczania na kurs przygotowawczy, płeć, czy dieta.

Poszukiwanie najlepszej konfiguracji sieci neuronowej

Przeprowadzono badanie błędu średniokwadratowego dla uczenia oraz testu różnych konfiguracji. Przetestowano konfigurację sieci o jednej oraz dwóch warstwach, z wykorzystaniem różnych kombinacji funkcji `logsig`, `tansig`, `purelin`, `radbas`. Dla wszystkich konfiguracji utworzono wykresy pudełkowe błędów, wnioski z przeprowadzonych badań zawarto w sprawozdaniu.

Najlepsza znaleziona konfiguracja została przedstawiona poniżej.



Rysunek 2: Przykładowe MSE uczenia dla 20 neuronów, 20 prób oraz kombinacji funkcji `purelin`, `purelin`.

Badanie wpływu ilości kolumn na jakość sieci

Dla najlepszej konfiguracji sieci testowano uczenie oraz działanie sieci bez kolejnych kolumn, w celu zbadania jaki wpływ mają one na działanie sieci.

W przypadku usunięcia kolumny, która informowała o wykształceniu rodziców otrzymaliśmy największe zmniejszenie błędów średniokwadratowych. Usunięcie kolumny, która świadczyła o przejściu kursu przygotowawczego spowodowało znaczne powiększenie błędów. W pozostałych przypadkach otrzymaliśmy podobne wyniki.

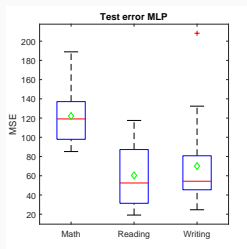
Może to świadczyć o dużym znaczeniu podjęcia się kursu przygotowawczego (polepsza działanie sieci) oraz wykształceniu rodziców (pogarsza działanie sieci) na wyniki egzaminów.

Uczenie sieci wynikami egzaminów

W ostatniej części projektu zostało przeprowadzone uczenie sieci, wykorzystując jako zbiór wejściowy wyniki z pozostałych egzaminów.

Sieć osiągała wtedy najlepsze rezultaty - błąd zmalał nawet czterokrotnie!

Związane jest to z faktem wysokiej korelacji pomiędzy wynikami egzaminów.



Rysunek 3: MSE testowania dla najlepszej konfiguracji sieci.

Podsumowanie

Z powodzeniem udało nam się znaleźć sieć neuronową o najmniejszej wartości błędu średniokwadratowego - jest to konfiguracja z liniowymi funkcjami aktywacji oraz 20 neuronami w warstwie ukrytej.

Choć wartości błędu średniokwadratowego są duże, gdy przeanalizujemy wyniki działania sieci, zauważymy, że dla poszczególnych wyników błąd jest najczęściej bliski zeru. Gdy dopuścimy margines błędu np 5 lub 10 punktów jakość wyników sieci znacząco wzrośnie.

Poszczególne kolumny mają wpływ na działanie sieci - pozytywny (kurs przygotowawczy) bądź negatywny (wykształcenie rodziców).

Sieć osiąga najlepsze rezultaty, gdy do uczenia użyjemy wyników pozostałych egzaminów, jednak nie są one zawsze dostępne.

Bibliografia



Assessment of artificial neural network for bathymetry estimation using high resolution satellite imagery in shallow lakes: Case study el burullus lake.

https://www.researchgate.net/publication/303875065_ASSESSMENT_OF_ARTIFICIAL_NEURAL_NETWORK_FOR_BATHYMETRY_ESTIMATION_USING_HIGH_RESOLUTION_SATELLITE_IMAGERY_IN_SHALLOW_LAKES_CASE_STUDY_EL_BURULLUS_LAKE.
[Online].



Kaggle.

<https://www.kaggle.com/spscientist/students-performance-in-exams>.
[Online].



Perceptron wielowarstwowy.

[https:](https://pl.wikipedia.org/wiki/Perceptron_wielowarstwowy)

[//pl.wikipedia.org/wiki/Perceptron_wielowarstwowy.](https://pl.wikipedia.org/wiki/Perceptron_wielowarstwowy)

[Online].



Sztuczne sieci neuronowe - wyklad.

[https:](https://upel2.cel.agh.edu.pl/wfiis/course/view.php?id=134)

[//upel2.cel.agh.edu.pl/wfiis/course/view.php?id=134.](https://upel2.cel.agh.edu.pl/wfiis/course/view.php?id=134)

[Online].