

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. 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* [1], 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 pisanía.

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:

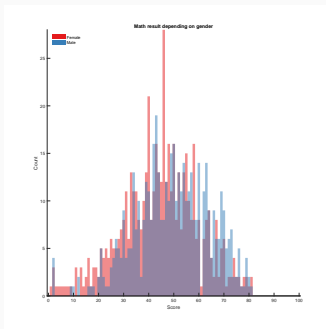
- Jednowarstwowe sieci neuronów dyskretnych
- Sieć perceptronów wielowarstwowych (ang. MLP)

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.

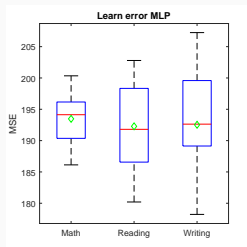


Rysunek 1: Przykład porównania wyników z matematyki w zależności od płci.

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, nie dało się jednak wysnuć wniosków dotyczących rasy, gdyż dane zostały nazwane tylko kolejnymi literami alfabetu.

Poszukiwanie najlepszej konfiguracji sieci neuronowej

Przeprowadzono badanie błędu średniokwadratowego dla uczenia oraz testu różnych konfiguracji funkcji aktywacji, ilości warstw ukrytych oraz ilości neuronów w poszczególnych warstwach. 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.



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

Badanie wpływu ilości kolumn na jakość sieci

Dla wybranej poprzednio konfiguracji najlepszych parametrów sieci zostało przeprowadzone uczenie ze zmniejszoną ilością kolumn. Kolejno 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 wykształcenia rodziców oraz podjęcia się kursu przygotowawczego na wynik egzaminu tych dwóch kolumn w pracy sieci.

Bibliografia



Students performance in exams.

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