

WYDZIAŁ FIZYKI I INFORMATYKI STOSOWANEJ



SZTUCZNE SIECI NEURONOWE
Sprawozdanie z projektu

Aleksandra Poręba
Grzegorz Podsiadło

14 czerwca 2020

Spis treści

1	Wstęp	3
2	Sieć neuronowa	4
3	Badanie wpływu ilości kolumn na wyniki sieci	5
3.1	Korelacja pomiędzy danymi	5
3.2	Testowanie sieci neuronowej	7
4	Przewidywanie wyniku egzaminu na podstawie pozostałych	8
5	Podsumowanie	9

1 Wstęp

2 Sieć neuronowa

3 Badanie wpływu ilości kolumn na wyniki sieci

3.1 Korelacja pomiędzy danymi

W ramach projektu został obliczony współczynnik korelacji *Pearsona* pomiędzy danymi z repozytorium. Dzięki niemu możemy sprawdzić jak bardzo zależne są od siebie dane, a dzięki temu dostosować informacje użyte do uczenia sieci.

Do obliczenia współczynnika korelacji została użyta funkcja pakietu MATLAB `corr`.

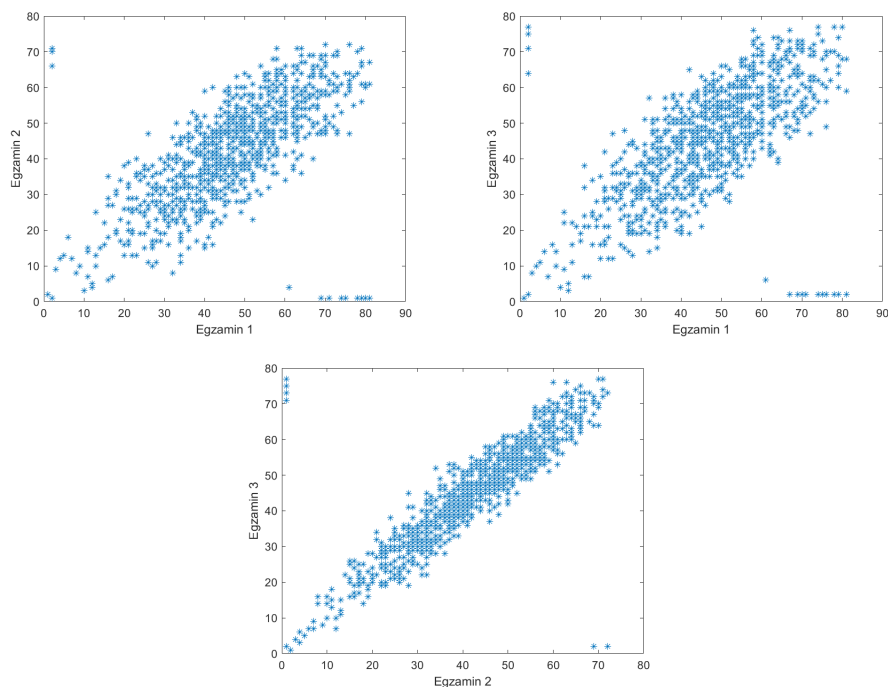
3.1.1 Korelacja pomiędzy wynikami egzaminów

Na początku został wyznaczony współczynnik korelacji pomiędzy wynikami egzaminów. Otrzymane wyniki przedstawiono w tabeli poniżej.

Egzamin 1 i 2	Egzamin 1 i 3	Egzamin 2 i 3
0.6453	0.6435	0.8578

Dla wszystkich kombinacji otrzymaliśmy wartości większe od 0.5, możemy więc uznać, że dane są od siebie zależne. Największa korelacja występuje pomiędzy egzaminem 2 oraz 3 - współczynnik jest równy 0.86.

Na rysunkach poniżej zależności pomiędzy zbiorami zostały przedstawione w sposób graficzny.



Rysunek 1: Zależności pomiędzy wynikami egzaminów przedstawione w sposób graficzny.

3.1.2 Korelacja pomiędzy czynnikami środowiskowymi a wynikami egzaminów

W dalszej części analizy zbioru danych została zbadana zależność pomiędzy czynnikami, będącymi wejściem sieci neuronowej, a wynikami kolejnych egzaminów. Otrzymane współczynniki korelacji przedstawiono w tabeli poniżej.

Czynnik środowiskowy	Egzamin 1	Egzamin 2	Egzamin 3
Płeć	0.1558	-0.1886	-0.2396
Rasa	0.1771	0.0770	0.0907
Wykształcenie rodzica	-0.0584	-0.0306	-0.0571
Przystąpienie do kursu	0.3269	0.1906	0.2182
Dieta	-0.1564	-0.1838	-0.2647

Otrzymane wartości są dość niskie, nie istnieje wyraźna korelacja pomiędzy którąś z tych cech, a wynikami. Najmniejszą zależność obserwujemy

po między wynikami, wykształceniem rodziców - są one najbliższe zeru. Największe znaczenie ma przystąpienie do kursu przygotowawczego, w dalszej kolejności dieta oraz płeć.

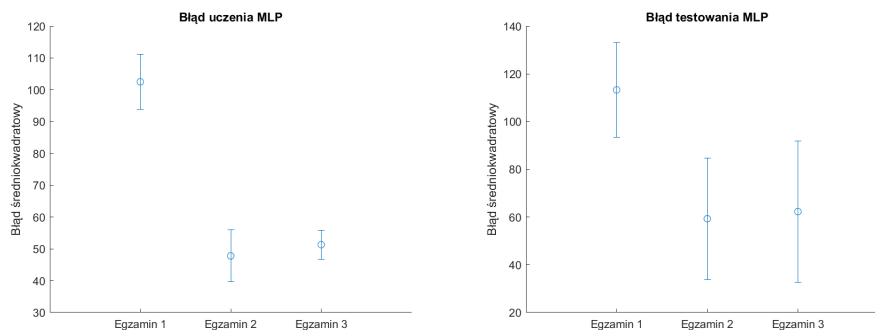
3.2 Testowanie sieci neuronowej

Dla wybranych najlepszych parametrów sieci zostało przeprowadzone uczenie z pomniejszoną ilością kolumn. Celem tego zabiegu było zbadanie, jaki wpływ mają te czynniki na poprawność działania sieci.

4 Przewidywanie wyniku egzaminu na podstawie pozostałych

Jak zostało zauważone podczas badania korelacji, wyniki egzaminów są od siebie zależne (wysoki współczynnik korelacji). Korzystając z wyników z dwóch egzaminów podjęto próbę stworzenia sieci obliczającej wynik z trzeciego egzaminu.

Otrzymane błędy testowania zostały przedstawione poniżej.



Rysunek 2: Błąd uczenia i testowania sieci dla X prób dla kolejnych egzaminów

Na podstawie obserwacji wielkości błędów, można zauważyć, że dla egzaminów 2 oraz 3 sieć osiąga dobre wyniki - na poziomie wartości 50. Dla egzaminu 1 błędy przyjmują wartości dwa razy większe.

Otrzymane wyniki można powiązać z analizowaną wcześniej korelacją danych - wyniki egzaminów 2 i 3 są ze sobą w większym stopniu powiązane, niż egzamin 1 z egzaminem 2 lub 3.

5 Podsumowanie

Literatura