

SIECI NEURONOWE – ćwiczenie 1

Celem ćwiczenia jest wprowadzenie/przypomnienie narzędzi i zapoznanie się z danymi z których będziemy korzystać w dalszej części kursu do ewaluacji sieci neuronowych jako metody uczenia maszynowego.

W tym i kolejnych ćwiczeniach – 2 i 3 - dozwolone jest korzystanie z dowolnych bibliotek do operowania na danych (wczytywanie, analiza eksploracyjna, preprocessing), uzyskania metryk jakości modelu i wydajnych obliczeń macierzowych. (Metody uczenia trzeba będzie jednak zaimplementować samodzielnie!) Sugerowana, prosta w użyciu konfiguracja to:

- Python
- Pandas do przeglądania danych
- Numpy do obliczeń na macierzach
- Scikit-learn do wyliczenia metryk jakości

Na pierwszych zajęciach masz czas na zapoznanie się z odpowiednimi bibliotekami/przypomnienie sobie ich z poprzedniego semestru. Następnie, wczytaj i przeanalizuj zbiór danych Heart Disease:

<https://archive.ics.uci.edu/dataset/45/heart+disease>

Analiza eksploracyjna powinna obejmować odpowiedzi na pytania:

- Czy zbiór jest zbalansowany pod względem liczby próbek na klasy?
- Jakie są średnie i odchylenia cech liczbowych?
- Dla cech liczbowych: czy ich rozkład jest w przybliżeniu normalny?
- Dla cech kategorycznych: czy rozkład jest w przybliżeniu równomierny?
- Czy występują cechy brakujące i jaką strategię możemy zastosować żeby je zastąpić?

W wyniku pracy do przyszłego tygodnia powinna powstać pierwsza część sumarycznego raportu który obejmować będzie ćwiczenia 1-4 oraz kod przekształcający dane do macierzy cech liczbowych (przykłady × cechy).

Ćwiczenie oceniane jest w skali 0-10 pkt.