

Wymagania: Realizacja projektu w którym podnosimy swój poziom wiedzy z zakresu statystycznego modelowania danych.

Cel: Zrozumienie realnie jak działa MSE za pomocą danych syntetycznych pochodzących z różnych rozkładów prawdopodobieństwa i modeli regresyjnych

Kryterium sukcesu: Każdy z członków zespołu osiąga epsilon progresu w obszarze jaki wspomniał na spotkaniu. Dla każdego z członków zespołu jest to inny obszar, a jako cały zespół chcemy głęboko zrozumieć jak działa MSE.

Epsilon progresu:

- **Diana Morzhak:** Analiza wpływu rozkładu danych na MSE jest uproszczoną analogią działania neuronaukowych modeli predictive coding, co umożliwia mi pogłębienie wiedzy o związkach pomiędzy uczeniem maszynowym a mechanizmami adaptacji mózgu
- **Mateusz Walo:** Zrozumienie podejścia do statystycznej analizy danych i uczenia maszynowego polegającego na analizie czynnika, którego nie mamy w danych, a mamy przypuszczenie, że ma wpływ na dane i na model, który chcemy zbudować. tzw causal approach in ML
- **Dominika Zydorczyk:** Rozwinięcie umiejętności interpretacji i komunikowania wyników analizy modeli regresyjnych. Lepsze zrozumienie, jak metryka MSE odzwierciedla jakość dopasowania modelu oraz jak różne typy danych i błędów wpływają na jej wartość

Opis:

Generujemy dane syntetyczne pochodzące z różnych rozkładów (testujemy różną liczbę zmiennych i różne rozkłady prawdopodobieństwa) i dokonujemy analizy statystycznej, budujemy i przeprowadzamy walidację modeli statystycznych starając się zrozumieć wpływ każdego z parametrów na wynik modelu i metrykę MSE. Badanie wpływu zmiennych, których "istnienia nie znamy" w ramce danych planujemy przez odpowiednie przygotowanie kilku zmiennych tak, aby miały "wspólną przyczynę".