

Instrukcja obsługi programu służącego do regresji liniowej:

Wymagania: System Windows 7 lub nowszy

Program jest aplikacją konsolową i wymaga czterech bądź pięciu (w zależności od algorytmu) argumentów. Pierwszym jest nazwa pliku.

Może się zdarzyć iż program nie wykonuje się od razu po uruchomieniu proszę w takim przypadku wcisnąć parokrotnie przycisk enter.

1. Plik źródłowy. Plik musi być w formacie csv z separatorem ','. Pierwsza linia zawiera nazwy argumentów. Ostatni argument to zawsze wartość funkcji dla podanych wcześniej argumentów. W kolejnych rzędach znajdują się liczby reprezentujące wartości argumentów i spodziewaną wartość funkcji. Wszystkie wartości po pierwszej linii muszą być liczbami w przeciwnym przypadku zachowanie programu może być nie zdefiniowane. Dla liczb rzeczywistych należy używać '.' Aby oddzielić część dziesiętną. Przykładowy plik źródłowy znajduje się w projekcie i nazywa się „input.csv”.

Plik z danymi powinien zostać zamknięty przed uruchomieniem programu inaczej działanie programu może być nie zdefiniowane.

2. Drugim argumentem jest wartość z przedziału [0-1] określająca podział danych na testowe i treningowe.
3. Wybierz czy dane mają zostać znormalizowane do wartości z przedziału [0-1]. Wpisz 'y' jeżeli mają zostać znormalizowane.
4. Wybierz implementację regresji liniowej. Należy wybrać '1', '2' lub '3'. W przypadku wybrania innego znaku program zostanie zamknięty i należy uruchomić go ponownie.
 - 1 regresja liniowa sumaryczna
 - 2 regresja liniowa alternatywna implementacja
 - 3 metoda gradientowa (iteracyjna)

4.b W przypadku wyboru algorytmu numer 3 należy wybrać ilość iteracji. Ilość iteracji musi być większa niż 0 i być liczbą całkowitą. Przy dużym zbiorze danych i dużej liczbie iteracji wykonanie programu może zająć znaczną ilość czasu.

5. Program wytrenuje model regresji i policzy MSE (Mean Squared Error), SSE(Sum of Squared Errors) oraz MAE (Mean absolute Error) dla zbioru treningowego i wypisze ich wartości. Następnie dla każdego zestawu argumentów z zbioru testowego wyznaczy obliczoną przez model spodziewaną wartość. Jako ostatni wypisze MSE, SSE oraz MAE dla zbioru testowego.

Następnie program będzie czekał na dowolny znak po czym zostanie zamknięty.. Aby uruchomić inny algorytm należy ponownie uruchomić program. Nie zaleca się zamykania programu w inny sposób gdyż może prowadzić do wycieków pamięci i uszkodzenia stosu.

Przykładowe uruchomienie programu zakładając że plik input.csv znajduje się w tym samym folderze co plik .exe

- ./MLWRegression.exe input.csv 0.3 n 3 1000
- ./MLWRegression.exe input.csv 0.15 y 2

Mikołaj Stoń