

METODY EKSPLOKACJI DANYCH

Laboratorium. Analiza regresji. Dobór zmiennych do modelu wieloczynnikowego Regresja z wykorzystaniem tzw. regularyzacji

WPROWADZENIE TEORETYCZNE

Patrz plik o nazwie „MED-lab-4-Regularyzacja-Wprowadzenie.pdf”

ZADANIE

Zadanie odnosi się do już znanego Państwu zagadnienie (zad.2) z laboratorium 1. odnoszącego się do poszukiwania zależności procentowej zawartości tłuszczu. Pozwolę zacytować sobie początkowa treść wspomnianego zadania

„Zebrano pomiary 250 mężczyzn w różnym wieku (źródło: <http://www.byu.edu/chhp>). Zebrane dane zawierają poprawnie zmierzony % tłuszczu „Pct.BF” (%bodyfat) oraz inne zweryfikowane wartości zmiennych charakteryzujących sylwetkę (...) Pełny zestaw danych do obecnego zadania znajduje się w pliku o nazwie „MED-lab-1-Zadanie-2-Dobor zmiennych-bodyfat-dane-i-opis.txt”...”

Proszę:

- 1) Zbudować model z wykorzystaniem techniki regularyzacji L1 lub L2, który pozwala przewidzieć „%bodyfat” na podstawie innych zmiennych.
- 2) Zbudować model z wykorzystaniem techniki regularyzacji L1 lub L2, który pozwala przewidzieć „density” na podstawie innych zmiennych.
- 3) Porównać rezultat z wcześniej (w lab.1) uzyskanymi wynikami.
- 4) Przygotować w krótkiej formie pisemnej sprawozdanie z wykonania zadania, które wraz z innymi plikami (wymaganymi do weryfikacji rozwiązania zadania) należy wstępnie umieścić w prywatnym notesie zespołu MS Teams w sekcji „Zaliczenie” w zakładce „Egzamin 0” – dodając zakładkę, jeżeli trzeba.