Artur Gęsiarz

Marzec 16, 2024

Laboratorium nr 2

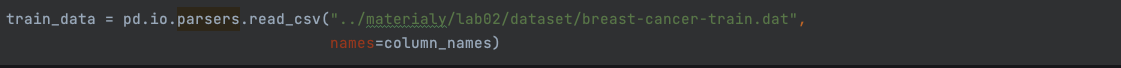
MOwNiT – Metoda najmniejszych kwadratów

1. Treść zadania
   1. Celem zadania jest zastosowanie metody najmniejszych kwadratów do predykcji, czy nowotwór jest złośliwy czy łagodny. Nowotwory złośliwe i łagodne mają różne charakterystyki wzrostu. Istotne cechy to m. in. promień i tekstura. Charakterystyki te wyznaczane są poprzez diagnostykę obrazową i biopsje.

Do rozwiązania problemu wykorzystamy bibliotekę ***pandas,*** typ ***DataFrame*** oraz dwa zbiory danych:

* *breast-cancer-train.dat*
* *breast-cancer-validate.dat.*

Nazwy kolumn znajdują się w pliku ***breast-cancer.labels***. Pierwsza kolumna to identyfikator pacjenta ***patient******ID****.* Dla każdego pacjenta wartość w kolumnie ***Malignant/Benign*** wskazuje klasę, tj. czy jego nowotwór jest złośliwy czy łagodny. Pozostałe 30 kolumn zawiera cechy, tj. charakterystyki nowotworu.

1. Rozwiązanie zadania
   1. Importowanie danych do testowania:  
        
      
   2. Implementacja rysowania histogramu charakterysytki promienia  
        
      Obraz zawierający tekst, zrzut ekranu, Czcionka

      Opis wygenerowany automatycznie
   3. Implementacja wykresu charakterystki promienia

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

* 1. Implementacja reprezentacji danych zawartych w obu zbiorach dla liniowej i kwadratowej metody najmniejszych kwadratów

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

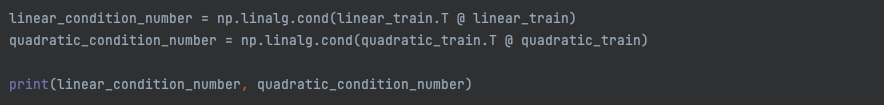
* 1. Implementacja wektora b dla obu zbiorów



* 1. Implementacja wagi dla liniowej oraz kwadratowej metody najmniejszych kwadratow



* 1. Implementacja obliczenia współczynnika uwarunkowania macierzy



* 1. Sprawdzenie jak dobrze otrzymane wagi przewidują typ nowotworu

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

* 1. Implementacja wykresu porównującego fałszywie dodatnie oraz fałszywie ujemne wyniki testu na nowotwór

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

1. Wykresy
   1. Histogram charakterystki promienia

Obraz zawierający diagram, Wykres, linia, zrzut ekranu

Opis wygenerowany automatycznie

**Wykres 1. Histogram opisujący jaki procent badanych ma jaki promień**

* 1. Wykres charakterystki promienia

Obraz zawierający tekst, zrzut ekranu, Wykres, diagram

Opis wygenerowany automatycznie

**Wykres 2. Wykres opisujący badanych a ich promień choroby**

* 1. Wykres porównujący fałszywie dodatnie testy oraz fałszywie ujemne otrzymane poprzez dwie różne metody

Obraz zawierający tekst, zrzut ekranu, diagram, Wykres

Opis wygenerowany automatycznie

**Wykres 3. Wykres porównujący metody oraz fałszywie dodatnie/ujemne wyniki testów**

1. Tabele
   1. Wyniki fałszywych dodanie oraz fałszywych ujemnie testów w poszczególnych metodach

|  |  |  |
| --- | --- | --- |
| Metoda | Fałszywie ujemne testy | Fałszywie dodanie testy |
| Liniowa | 2 | 6 |
| Kwadratowa | 7 | 8 |

**Tabela 1. Porównanie wyników testów**

* 1. Tabela współczynników uwarunkowania macierzy

|  |  |
| --- | --- |
| Metoda | Współczynnik uwarunkowania macierzy |
| Liniowa |  |
| Kwadratowa |  |

1. Wnioski

Analizując wyniki, można zauważyć, że zarówno liniowa, jak i kwadratowa metoda najmniejszych kwadratów miały podobne wyniki w przypadku fałszywie dodanich, ale miały różne wyniki w przypadku fałszywie ujemnych.

Mimo wszystko lepiej poradziła sobie metoda liniowa poniewaz ma mniej przypadkow falszywie ujemnych od metody kwadratowej.

Jednak porównując współczynniki uwarunkowania macierzy znacznie większy ma metoda liniowa od kwadratowej.

Podsumowując, różnice w wynikach między liniową a kwadratową metodą najmniejszych kwadratów mogą wynikać z różnych czynników, takich jak złożoność modelu, rozmiar zestawu treningowego, wybór cech i zakłócenia w danych.

1. Bibliografia

*Wykład MOwNiT - prowadzony przez dr. Inż. K. Rycerz  
Prezentacje – dr. Inż. M. Kuta*

1. Dodatkowe informacje

Rozwiązania zadania znajduje się w pliku ex1.ipynb