

Metoda najmniejszych kwadratów

Laboratorium 2

Jakub Ciszewski, Wiktor Smaga

21 marca 2024

1 Cel ćwiczenia

Zapoznanie się z metodą najmniejszych kwadratów oraz zastosowanie jej w praktyce.

2 Opis ćwiczenia

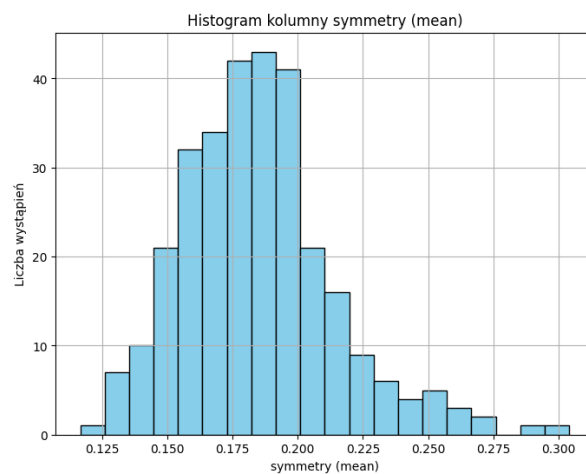
Ćwiczenie polegało na wczytaniu odpowiednich danych dotyczących nowotworów, zwizualizowanie ich oraz użycie liniowej i kwadratowej metody najmniejszych kwadratów do przewidzenia na podstawie wybranych parametrów złożoności nowotworu.

3 Dane

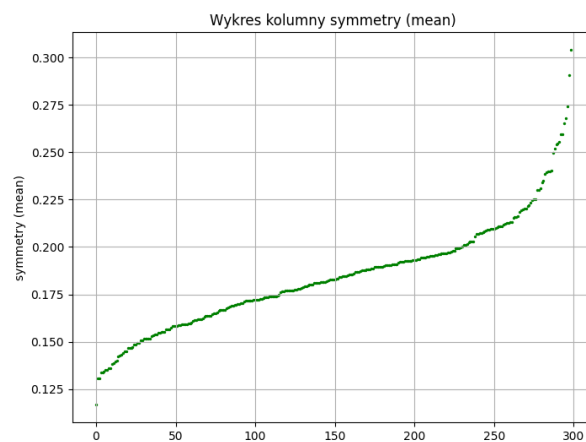
Wczytano dwa zestawy danych

- Zestaw `train` przeznaczony do trenowania modelu.
- Zestaw `validate` przeznaczony do sprawdzenia wytrenowanego modelu.

4 Wizualizacja wybranych danych



Wykres 1: Ilości wystąpień nowotworu złośliwego dla średnich symetrii w zbiorze danych `train`.



Wykres 2: Średnie symetrie w zbiorze danych `train`.

5 Metoda najmniejszych kwadratów

Wszystkie macierze są tworzone dla obu zestawów `train` oraz `validate`.

5.1 Macierz liniowa

Do stworzenia macierzy liniowej użyto wszystkich danych.

$$X_{lin} = \begin{bmatrix} f_{1;1} & f_{1;2} & \cdots & f_{1;30} \\ f_{2;1} & f_{2;2} & \cdots & f_{2;30} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n;1} & f_{n;2} & \cdots & f_{n;30} \end{bmatrix}$$

$$Y = \begin{bmatrix} g_{1;1} \\ g_{2;1} \\ \vdots \\ g_{n;1} \end{bmatrix}, \quad g_{i;j} = \begin{cases} 1 & \text{jeśli nowotwór } i \text{ złośliwy} \\ -1 & \text{jeśli nowotwór } i \text{ łagodny} \end{cases}$$

5.2 Macierz kwadratowa

Do stworzenia macierzy kwadratowej użyto tylko 4 kolumn danych z zestawu: **radius (mean), perimeter (mean), area (mean), symmetry (mean)**.

$$X_{quad} = \begin{bmatrix} f_{1;1} & \cdots & f_{1;4} & f_{1;4}^2 & \cdots & f_{1;4}^2 & f_{1;1}f_{1;2} & \cdots & f_{1;3}f_{1;4} \\ f_{2;1} & \cdots & f_{2;4} & f_{2;4}^2 & \cdots & f_{2;4}^2 & f_{2;1}f_{2;2} & \cdots & f_{2;3}f_{2;4} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{n;1} & \cdots & f_{n;4} & f_{n;4}^2 & \cdots & f_{n;4}^2 & f_{n;1}f_{n;2} & \cdots & f_{n;3}f_{n;4} \end{bmatrix}$$

Macierz Y jest dokładnie taka sama jak w wersji liniowej.

5.3 Macierz wag

Macierze wag dla macierzy liniowej i kwadratowej uzyskano w ten sam sposób, czyli rozwiązując równanie:

$$w = (A^T A)^{-1} A^T b$$

w - macierz wag

A - macierz liniowa/kwadratowa

b - macierz Y

5.4 Macierz predykcji

Utworzono macierz P zawierającą predykcje złośliwości nowotworów mnożąc macierz wag z odpowiednią macierzą liniową/kwadratową zbioru `validate`.

$$P_{lin} = X_{lin} * w$$

$$P_{quad} = X_{quad} * w$$

Na podstawie tych macierzy skategoryzowano dane:

- $P_i > 0 \wedge Y_i = -1$ - wynik fałszywie dodatni
- $P_i \leq 0 \wedge Y_i = 1$ - wynik fałszywie ujemny

6 Wyniki

6.1 Trafność predykcji

	Fałszywie dodatnie	Fałszywie ujemne	Błąd względny
Macierz liniowa	15	5	96.91%
Macierz kwadratowa	6	2	92.28%

Tabela 1: Zestawienie wyników predykcji dla reprezentacji kwadratowej i liniowej zbioru `train`

6.2 Współczynnik uwarunkowania

Obliczono współczynniki uwarunkowania dla macierzy kwadratowej i liniowej zbioru `train` korzystając z funkcji `cond` z biblioteki `numpy`. Wykorzystuje ona metodę **SVD** (*Singular Value Decomposition*).

	współczynnik uwarunkowania
macierz kwadratowa	9.028541478152603e+17
macierz liniowa	1803175716260.0334

Tabela 2: Współczynniki uwarunkowania macierzy kwadratowej i liniowej dla zbioru `train`

7 Wnioski

- Problem jest bardzo źle uwarunkowany.
- Im więcej danych tym większa dokładność.
- Reprezentacja kwadratowa pozwala nam zwiększyć precyzję przewidywań przy ograniczonym zasobie danych.

8 Bibliografia

- https://pl.wikipedia.org/wiki/Wska%C5%BAnik_uwarunkowania
- https://pl.wikipedia.org/wiki/Rozk%C5%82ad_wed%C5%82ug_warto%C5%9Bci_osobliwych
- https://pl.wikipedia.org/wiki/Metoda_najmniejszych_kwadrat%C3%B3w
- http://heath.cs.illinois.edu/scicomp/notes/cs450_chapt03.pdf