

Laboratorium 02 - Metoda najmniejszych kwadratów

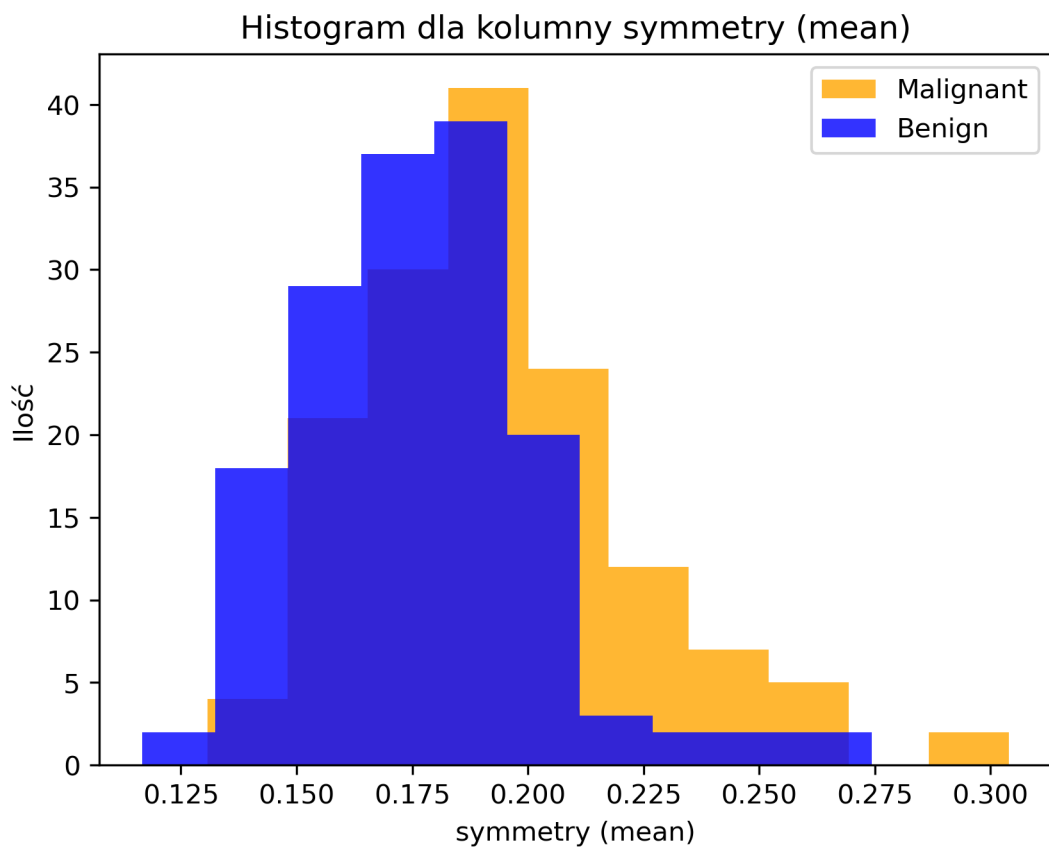
Błażej Naziemiec i Szymon Żuk

18 marca 2025

Wstęp

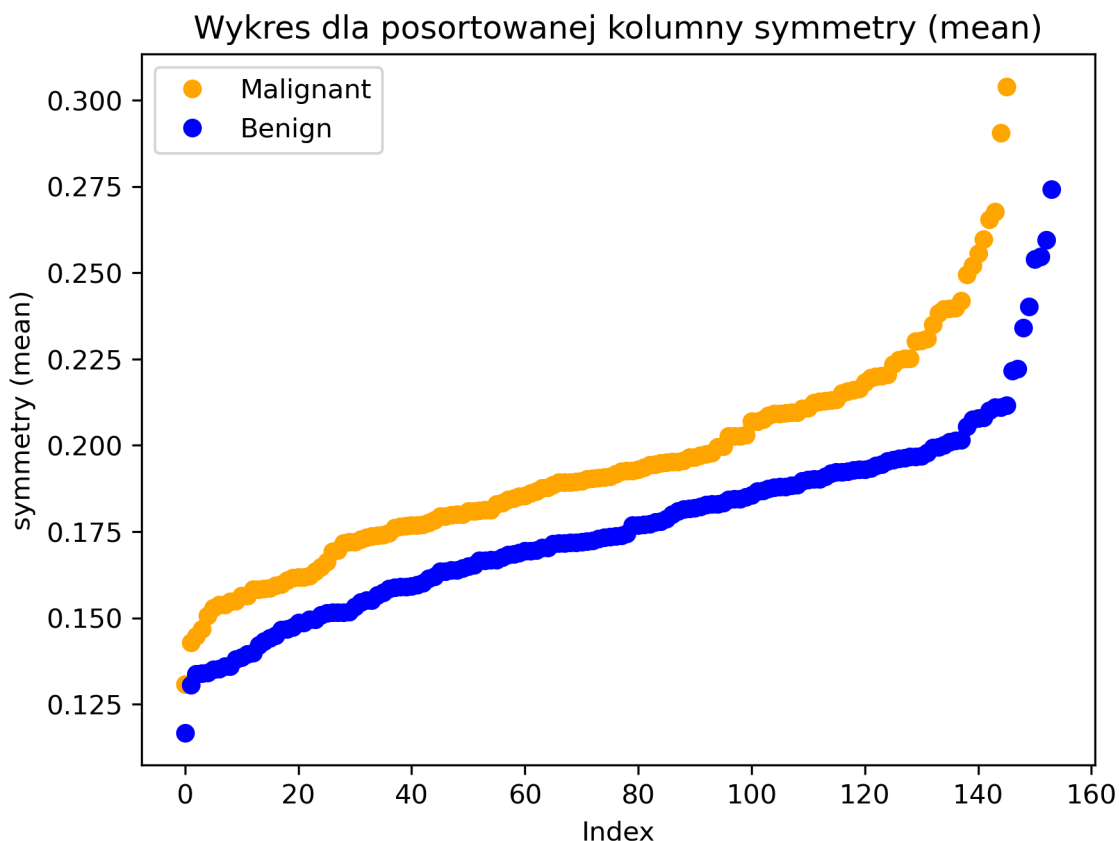
Celem zadania jest zastosowanie metody najmniejszych kwadratów do predykcji, czy nowotwór jest złośliwy (ang. *malignant*) czy łagodny (ang. *benign*). Nowotwory złośliwe i łagodne mają różne charakterystyki wzrostu. Istotne cechy to m. in. promień i tekstura. Charakterystyki te wyznaczone są poprzez diagnostykę obrazową i biopsje.

Na początku zaczytaliśmy dane z przygotowanych plików .dat i wrzuciliśmy je na histogram oraz wykres punktowy. W wykresie punktowym zgodnie z poleceniem posortowaliśmy wartości kolumny od najmniejszej do największej. Wybraliśmy *symmetry (mean)* do przedstawienia na histogramie oraz wykresie.



Rysunek 1. Histogram dla cechy symmetry (mean)

Rozkład dla nowotworów złośliwych jest przesunięty w prawo, co wskazuje, że złośliwe guzy nowotworowe zazwyczaj są bardziej symetryczne niż łagodne. Największa koncentracja dla nowotworów złośliwych jest w okolicach 0.200, a dla łagodnych w okolicach 0.180.



Rysunek 2 Wykres punktowy dla cechy symmetry (mean)

Dane dla nowotworów złośliwych (pomarańczowe punkty) są generalnie wyżej na wykresie niż dane dla nowotworów łagodnych, co potwierdza wcześniejszą obserwację, że złośliwe guzy nowotorowe są bardziej symetryczne niż łagodne.

Macierze dla liniowej i kwadratowej metody najmniejszych kwadratów

Przygotowanie danych

Stworzyliśmy reprezentacje danych zawartych w obu zbiorach dla liniowej i kwadratowej metody najmniejszych kwadratów. W następnym kroku stworzyliśmy wektor b dla zbiorów nowotworów łagodnych oraz złośliwych. Następnie znaleźliśmy wagi dla liniowej oraz kwadratowej reprezentacji najmniejszych kwadratów przy pomocy wcześniej stworzonych macierzy i wektora b z zbioru danych treningowych. Znaleźliśmy też zbiór wag przy użyciu funkcji `np.linalg.lstsq` z

$$\lambda = 0.01$$

Wyznaczenie współczynnika uwarunkowania macierzy dla liniowej i kwadratowej metody najmniejszych kwadratów

Za pomocą funkcji `np.linalg.cond` wyznaczyliśmy współczynnik uwarunkowania macierzy dla liniowej i kwadratowej metody najmniejszych kwadratów. Dla poszczególnych metod otrzymaliśmy następujące wyniki:

- Dla liniowej metody najmniejszych kwadratów: $1.8092 * 10^{12}$
- Dla kwadratowej metody najmniejszych kwadratów: $9.0568 * 10^{17}$

Wartości zarówno dla liniowej, jak i kwadratowej metody najmniejszych kwadratów są duże, co oznacza, że macierze są źle uwarunkowane. Wartości dla macierzy kwadratowej są bardzo wysokie.

Sprawdzenie, jak dobrze otrzymane wagi przewidują typ nowotworu

Dla poszczególnych metod przedstawiamy wyniki klasyfikacji dla danych testowych. W tabeli przedstawiamy wartości TP, TN, FP, FN oraz dokładność klasyfikacji.

- TP - liczba przypadków prawdziwie dodatnich
- TN - liczba przypadków prawdziwie ujemnych
- FP - liczba przypadków fałszywie dodatnich
- FN - liczba przypadków fałszywie ujemnych
- Accuracy - dokładność klasyfikacji ($\frac{TP+TN}{TP+TN+FP+FN}$)

TP	TN	FP	FN	Accuracy
58	194	6	2	96.92%

Tabela 1. Wyniki klasyfikacji dla liniowej metody najmniejszych kwadratów

TP	TN	FP	FN	Accuracy
55	185	15	5	92.31%

Tabela 2. Wyniki klasyfikacji dla kwadratowej metody najmniejszych kwadratów

Porównując wyniki przedstawione w tabelach (1) i (2) można zauważyć, że liniowa metoda najmniejszych kwadratów daje lepsze wyniki klasyfikacji niż kwadratowa metoda najmniejszych kwadratów. Dokładność klasyfikacji dla liniowej metody wynosi 96.92%, a dla kwadratowej 92.31%. Warto jednak zauważyć, że poprawność wyników obu metod jest bardzo wysoka. Świadczy to o poprawności zaimplementowanych rozwiązań.