

# Lab02 Metoda najmniejszych kwadratów

Patryk Blacha, Radosław Szepielak

18 czerwca 2025

## Spis treści

<b>1</b>	<b>Wprowadzenie teoretyczne</b>	<b>1</b>
1.1	Metoda najmniejszych kwadratów w klasyfikacji . . . . .	1
<b>2</b>	<b>Metodologia</b>	<b>1</b>
2.1	Zbiór danych i przygotowanie . . . . .	1
2.2	Proces przetwarzania . . . . .	2
<b>3</b>	<b>Analiza numeryczna</b>	<b>2</b>
3.1	Współczynniki uwarunkowania . . . . .	2
<b>4</b>	<b>Wyniki eksperymentalne</b>	<b>3</b>
4.1	Porównanie skuteczności . . . . .	3
4.2	Szczegółowa analiza błędów . . . . .	3
<b>5</b>	<b>Dyskusja</b>	<b>3</b>
5.1	Interpretacja wyników . . . . .	3
5.2	Ograniczenia metody . . . . .	3

## 1 Wprowadzenie teoretyczne

### 1.1 Metoda najmniejszych kwadratów w klasyfikacji

Metoda najmniejszych kwadratów (MNK) została zastosowana do problemu klasyfikacji binarnej poprzez odpowiednie przekształcenie etykiet. Dla macierzy cech  $\mathbf{A}$  i wektora  $\mathbf{b}$  rozwiązanie wyraża się wzorem:

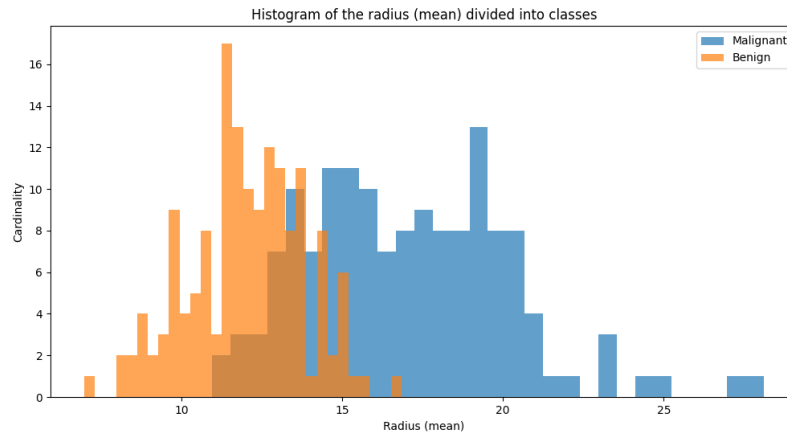
$$\mathbf{w} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$

gdzie  $\mathbf{w}$  zawiera optymalne wagi modelu.

## 2 Metodologia

### 2.1 Zbiór danych i przygotowanie

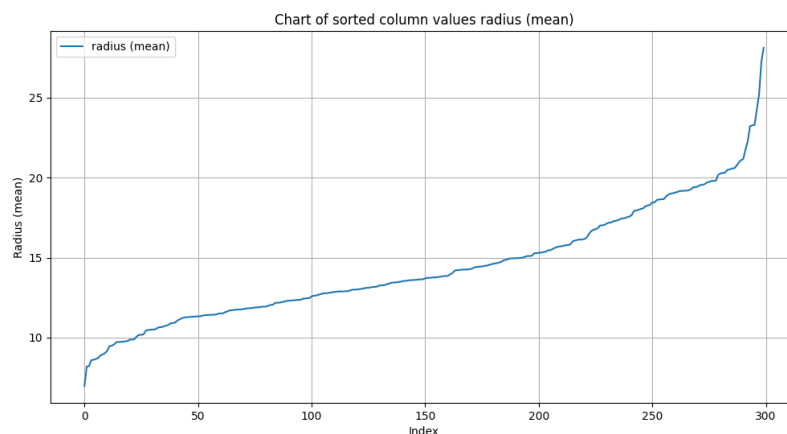
- Zbiór Breast Cancer Wisconsin zawierający 30 cech diagnostycznych
- Podział na treningowy (300 próbek) i walidacyjny (260 próbek)
- Kodowanie etykiet: 1 (złośliwy), -1 (łagodny)



Rysunek 1: Rozkład wartości cechy 'radius (mean)' z podziałem na klasy

## 2.2 Proces przetwarzania

1. Normalizacja brakująca (dane już wystandaryzowane)
2. Generowanie cech kwadratowych dla wybranych atrybutów
3. Budowa macierzy projektowych dla modeli liniowych i nieliniowych



Rysunek 2: Posortowane wartości cechy 'radius (mean)'

## 3 Analiza numeryczna

### 3.1 Współczynniki uwarunkowania

- Współczynnik uwarunkowania dla modelu liniowego:  $2.10 \times 10^{12}$   
Ten współczynnik jest wysoki, co wskazuje na potencjalne problemy numeryczne. Macierz A jest prawdopodobnie źle uwarunkowana, co może prowadzić do niestabilnych wyników.
- Współczynnik uwarunkowania dla modelu kwadratowego:  $1.07 \times 10^{18}$   
Ten współczynnik jest dużo wyższy niż dla modelu liniowego, co wskazuje na złe uwarunkowanie macierzy. Wyniki uzyskane tą metodą są prawdopodobnie mało wiarygodne.
- Współczynnik uwarunkowania dla modelu liniowego z regularyzacją:  $5.31 \times 10^{10}$   
Współczynnik uwarunkowania jest niższy niż w przypadku metody bez regularyzacji i metody

kwadratowej, ale nadal wysoki. Regularyzacja poprawiła uwarunkowanie macierzy, ale nadal istnieje ryzyko niestabilności.

- Współczynnik uwarunkowania dla modelu liniowego SVD:  $1.45 \times 10^6$   
Współczynnik uwarunkowania jest znacznie niższy niż w poprzednich przypadkach, co sugeruje, że metoda SVD jest bardziej stabilna numerycznie.

## 4 Wyniki eksperymentalne

### 4.1 Porównanie skuteczności

Model	Dokładność
Liniowy	97.31%
Liniowy SVD	97.31%
Liniowy z regularyzacją	97.69%
Kwadratowy	91.92%

Tabela 1: Porównanie efektywności modeli

### 4.2 Szczegółowa analiza błędów

Model	TP	TN	FP	FN
Liniowy	58	195	5	2
Liniowy SVD	58	195	5	2
Liniowy z regularyzacją	58	196	4	2
Kwadratowy	55	184	16	5

Tabela 2: Macierz pomyłek dla różnych modeli.

## 5 Dyskusja

### 5.1 Interpretacja wyników

- Wysoka dokładność modelu liniowego sugeruje liniową separowalność danych
- Spadek dokładności dla modelu kwadratowego wskazuje na:
  - Przeuczenie na szumach
  - Problemy numeryczne związane z wysokim współczynnikiem uwarunkowania

### 5.2 Ograniczenia metody

- Wrażliwość na współliniowość cech
- Brak automatycznej selekcji cech
- Problemy z interpretacją wag przy wysokiej korelacji zmiennych

### Pełne parametry modeli

```
1 Wagi modelu liniowego (pierwsze 5):  
2 - Bias: -6.172  
3 - radius_mean: -0.608  
4 - texture_mean: 0.025  
5 - perimeter_mean: 0.078  
6 - area_mean: 0.00058
```

## Literatura

- Materiały pomocnicze do laboratorium zamieszczone na platformie Teams (lab02/lab2-intro.pdf)