

# Regresja Liniowa

Uczenie Maszynowe  
uczeniemaszynowe@cs.uni.wroc.pl

21.10.2025, 10:00  
pracownia

## Uwagi ogólne:

- Wszystkie dane dostarczone są w plikach `.csv`.
- Zalecamy, żeby przy rozwiązywaniu zadań nie kopiować całości rozwiązań z modeli językowych. Można jednak zapytać o feedback i dodatkowe materiały na zadany temat.
- Podczas oddawania zadań weryfikacji poddane zostanie zrozumienie poszczególnych etapów zadania i znajomość elementów rozwiązania. Należy mieć przygotowanie środowisko z dostępem do wszystkich danych, tak by móc sprawnie zaimplementować drobne modyfikacje rozwiązań.
- Nie ma konieczności szczegółowo zapisywać odpowiedzi na pytania do dyskusji na pracowni, ale zdecydowanie warto te kwestie przemyśleć – choćby w tramwaju w drodze na zajęcia.

## Zadanie 1. [2 pkt]

Czy regresja liniowa jest modelem odpowiednim dla danych `data1.csv`? Jak odpowiednio przygotować dane do modelowania?

1 pkt. Zrób odpowiednie obliczenia i wizualizacje. Zapisz obserwacje i wnioski.

1 pkt. Zaproponuj jak przygotować dane lub przekształcić formułę, by móc do nich użyć modelu regresji liniowej. Rozwiązanie podaj w postaci jednej funkcji `preprocess_data(X_train, y_train)`, której działanie powinno być niezmiennicze na permutacje obserwacji w zbiorze uczącym.

## Wskazówki:

- Aby wczytać dane możesz użyć funkcji `pd.read_csv()` z biblioteki `pandas`, zwróć uwagę na nagłówki pliku.
- Wykresy możesz zrobić korzystając z bibliotek `matplotlib` lub `seaborn`.
- Żeby dobrać odpowiedni rodzaj wykresu postaw pytanie, na które chcesz poznać odpowiedź. Np. przypomnij sobie jakie są założenia o danych dla modelu regresji liniowej i sprawdź, czy dane spełniają te założenia. Wybierz jakie zmienne zwizualizować na osiach wykresu. Następnie, wybierz typ wykresu (`lineplot`, `scatterplot`, `histogram`, `KDEplot`, ...).
- Jakielikolwiek obliczenia na pracowni robimy na komputerze, preferowane jest użycie funkcji dostępnych w popularnych bibliotekach.
- Zacznij od opisanie co widzisz na wykresie (np. punkty reprezentujące dane układają się w okrąg), czyli obserwacji. Następnie przejdź do wniosków (np. zatem z dużym prawdopodobieństwem, zależność między zmienną na osi X oraz zmienną na osi Y nie jest funkcją).
- Spróbuj zidentyfikować prostą zależność między zmiennymi na osiach X i Y, zapisz równanie tej zależności. Pomoże Ci to zidentyfikować pełne rozwiązanie drugiej części zadania.

**Kryteria oceny lub wzorcowe rozwiązanie: TBA**

## Zadanie 2. [3 pkt]

Korzystając z biblioteki `sklearn`, dopasuj do danych `data1.csv` model regresji liniowej oraz sprawdź jego skuteczność, korzystając z błędu średniokwadratowego. Sprawdź, jak użycie przygotowania danych zaproponowanego w Zadaniu 1 wpływa na otrzymane wyniki.

- .5 pkt Zaproponuj odpowiedni podział danych. Czy będzie on taki sam przy budowanie pojedynczego modelu jak również w sytuacji, kiedy chcemy wybrać jeden spośród kilku modeli?
- .5 pkt Zaproponuj model baseline'owy, czyli prostszy model, który posłuży nam jako punkt odniesienia, czy uzyskane wyniki są akceptowalnej jakości.
- .5 pkt Zbuduj model regresji liniowej i model baseline'nowy na rozpatrywanych wariantach danych (po jednym dla surowych danych i danych po preprocessingu).
- .5 pkt Przeprowadź ewaluację modeli. Wyniki przedstaw w formie tabeli.
- 1 pkt Przeprowadź ewaluację modelu na danych udostępnionych w trakcie pracowni. *Ocena tego podpunktu będzie zależała od uzyskanych wyników.*

**Kryteria oceny lub wzorcowe rozwiązanie: TBA**

## Zadanie 3. [2 pkt]

Użyj danych `advertising.csv`.

- .5 pkt Przeprowadź analizę i przygotowanie danych, wzorując się na zadaniach 1 i 2.
- .5 kpt "Zbuduj modele regresji liniowej na danych, w których skala zmiennych jest zachowana i na danych znormalizowanych.
- .5 pkt Porównaj te dwa modele, zapisz obserwacje i wnioski.
- .5 pkt Wskaż najlepszy model do ewaluacji na zbiorze testowym udostępnionym w trakcie ćwiczeń.

**Wskazówka:** użyj doświadczenia (i kodów) z poprzednich zadań. Unikaj powtarzających się fragmentów kodu.

**Kryteria oceny lub wzorcowe rozwiązanie: TBA**

## Zadanie 4. [3 pkt]

Użyj danych `data4.csv`.

- .5 pkt Przeanalizuj dostarczone dane.
- 1 pkt Dla podanych danych zbuduj dwa modele: standardową regresję liniową oraz z użyciem regularyzacji w normie  $L_2$  z **odpowiednim** hiperparametrem  $\lambda_2$ .
- 1 pkt Porównaj otrzymane parametry regresji oraz skuteczność na zbiorze testowym. Który model zadziałał lepiej i dlaczego (podać obserwacje i teoretyczne uzasadnienie)?
- .5 pkt Sprawdź skuteczność modelu na udostępnionym w trakcie zajęć zbiorze danych.

**Kryteria oceny lub wzorcowe rozwiązanie: TBA**

## Dyskusja na pracownię

Jakie techniki i dobre praktyki poznaliście / zastosowaliście rozwiązując tę listę? Które z nich dotyczą regresji liniowej, a które będą miały przełożenie także na inne tematy/działy uczenia maszynowego?