

Zadania do wykonania na komputerach

Regresja jednowymiarowa i wielowymiarowa

Zadanie 1 (regresja jednowymiarowa) — sprzedaż lodów a temperatura

Zadanie 1. Kontekst. Badamy zależność sprzedaży lodów (liczba gałek sprzedanych dziennie) od średniej dziennej temperatury powietrza (w stopniach Celsjusza). Chcemy zbudować prosty model liniowy, który umożliwi przewidywanie sprzedaży na podstawie temperatury.

Dane. Wygeneruj zbiór 100 obserwacji według schematu:

$$sales = 5.0 \cdot temp + 50 + \varepsilon.$$

Liczbę ε losujemy jako liczbę całkowitą z zakresu od -10 do 10 . Temperaturę też losujemy jako liczbę całkowitą z zakresu od 0 do 30 .

Zadanie (kroki do wykonania).

1. Wczytaj lub wygeneruj dane i wyświetl pierwsze 10 wierszy.
2. Podziel dane na zbiór treningowy (80%) i testowy (20%).
3. Dopasuj model regresji liniowej (OLS): $sales \sim temp$.
4. Oblicz MSE i R^2 na zbiorze testowym.
5. Narysuj wykres punktowy ($temp$, $sales$) oraz wykreśl prostą dopasowaną na zbiorze treningowym.
6. Krótko zinterpretuj wyniki: co oznacza współczynnik kierunkowy (slope), intercept i czy model jest sensowny w kontekście sprzedaży lodów.

Do oddania.

- Pliki z danymi, programem oraz wynikami

Zadanie 2 (regresja wielowymiarowa) — przewidywanie zużycia energii w domu

Zadanie 2. Kontekst. Celem jest przewidzenie miesięcznego zużycia energii elektrycznej (kWh) w gospodarstwie domowym na podstawie cech budynku i użytkowania.

Dane. Wygeneruj zbiór 300 próbek z następującymi kolumnami:

- $area$ — powierzchnia mieszkania (m^2), losuj z przedziału $[40,200]$,
- $people$ — liczba mieszkańców, losuj z dyskretnego zbioru $\{1,2,3,4,5\}$,

- *windows* — liczba okien, losuj z $\{1, \dots, 10\}$,
- *age* — wiek budynku w latach, losuj z $[0, 80]$,
- *consumption* — miesięczne zużycie energii (kWh), generowane według:

$$consumption = 0.8 \cdot area + 10 \cdot people + 2 \cdot windows - 0.1 \cdot age + \varepsilon,$$

gdzie $\varepsilon \sim \mathcal{N}(0, 20^2)$.

Zadanie

1. Wczytaj lub wygeneruj dane i zapisz plik *energy.csv*.
2. Podziel dane na zbiór treningowy (80%) i testowy (20%).
3. Dopasuj regresję liniową wielowymiarową (OLS) do przewidywania *consumption*.
4. Oblicz MSE i R^2 na zbiorze testowym.
5. Wypisz współczynniki modelu i zinterpretuj ich znaczenie (jak zmiana każdej cechy wpływa na zużycie energii).
6. Przeprowadź standaryzację cech ($\text{srednia}=0$, $\text{wariancja}=1$) i dopasuj model jeszcze raz — porównaj współczynniki (w kontekście standardowych jednostek).

Do oddania:

- Plik *energy.csv* z wygenerowanymi danymi, pliki z programami oraz wynikami.