Regresja Liniowa

KNSM "Gauss"

Lato 2025

Ogólna postać modelu liniowego

Model liniowy przewiduje wartość poprzez obliczenie ważonej sumy cech wejściowych oraz składnika wolnego:

$$\hat{y}^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \cdots + \theta_n x_n^{(i)}.$$

- $\hat{y}^{(i)}$: Przewidywana wartość dla i-tej instancji.
- n: Liczba cech.
- $x_j^{(i)}$: Wartość j-tej cechy dla i-tej instancji.
- θ_j: Parametry modelu (w tym składnik wolny θ_0 oraz wagi $\theta_1, \theta_2, \ldots, \theta_n$).
- Zakładamy,że $\forall_{j=1,2,\ldots,m} \; x_0^{(j)} = 1$.

Postać wektorowa

Przewidywanie dla i-tej instancji można zapisać w zwartej postaci wektorowej:

$$\hat{y}^{(i)} = \theta \cdot x^{(i)}.$$

- θ: Wektor parametrów (składnik wolny + wagi).
- $x^{(i)}$: Wektor cech dla i-tej instancji (z $x_0^{(i)} = 1$ dla składnika wolnego).
- $\theta \cdot x^{(i)}$: Iloczyn skalarny wektorów θ i $x^{(i)}$.

Notacja macierzowa

Jeśli θ jest wektorem i X jest macierzą:

$$\hat{y} = X\theta.$$

- $\hat{y}$: Wektor przewidywań dla wszystkich instancji.
- X: Macierz cech (każdy wiersz odpowiada instancji).
- θ: Wektor parametrów (składnik wolny + wagi).

Uczenie modelu

Uczenie polega na znalezieniu parametrów θ, które minimalizują błąd predykcji.

- Miara wydajności: **Błąd Średniokwadratowy (MSE)**

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2.$$

- m: Liczba próbek.
- $\hat{y}^{(i)}$: Przewidywanie dla i-tej instancji.
- $y^{(i)}$: Rzeczywista wartość dla i-tej instancji.
- Minimalizowanie MSE minimalizuje również RMSE.

Minimalizacja MSE

Aby znaleźć optymalne parametry $\hat{\theta}$, minimalizujemy MSE:

- Używamy **Równania Normalnego** (rozwiązanie analityczne):

$$\hat{\theta} = (X^T X)^{-1} X^T y.$$

- Używanie algorytmów optymalizacji (np. gradientu prostego) jest tutaj nieoptymalne, wrór analityczny jest szybszy.

Wyprowadzenie Równania Normalnego

Minimalizujemy MSE w postaci macierzowej:

$$\mathrm{MSE} = \frac{1}{m}(X\theta - y)^T(X\theta - y).$$

Rozwijamy wyrażenie:

$$\mathrm{MSE} = \frac{1}{m}\left[\theta^T X^T X\theta - 2\theta^T X^T y + y^T y\right].$$

Pochodna względem θ:

$$\frac{\partial \mathrm{MSE}}{\partial \theta} = \frac{2}{m}(X^T X\theta - X^T y) = 0.$$

Rozwiązanie dla θ:

$$X^T X\theta = X^T y \implies \hat{\theta} = (X^T X)^{-1} X^T y. \quad \square$$