

KLASYCZNY MODEL REGRESJI LINIOWEJ

Zmienna objaśniana y w modelu regresji liniowej może zależeć od więcej niż jednej zmiennej objaśniającej x .

PRZYKŁAD

Wydajność reakcji chemicznej zależy od temperatury i stężenia katalizatora. Model wielowymiarowej regresji liniowej może opisać tę zależność jako

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

gdzie y oznacza wydajność, x_1 oznacza temperaturę i x_2 stężenie katalizatora.

W tym przypadku mamy dwa regresory a zależność jest liniową, funkcją nieznaną parametrów β_0 , β_1 i β_2 .

Tak więc model wielowymiarowej regresji liniowej opisujący zależność liniową zmiennej objaśnianej y z wykorzystaniem k regresorów jako zmienne objaśniające ma postać

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Model regresji liniowej jest zwykle wykorzystywany jako empiryczny model. Prawdziwa zależność y od $x_1, x_2, \dots, x_k$ jest zwykle nieznana jednak dla pewnego zakresu zmienności regresorów regresja liniowa jest zadawalającą aproksymacją prawdziwej nieznanej funkcji.

Model regresji liniowej można wykorzystywać do analizy modeli o bardziej złożonej zależności między regresorami i zmienną objaśnianą. Na przykład rozważmy model postaci

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon.$$

Jeśli oznaczymy przez $x_1 = x$, $x_2 = x^2$ i $x_3 = x^3$ to model ten przyjmuje postać

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon,$$

który jest modelem wielomianowej regresji liniowej.

Możemy również rozważyć interakcje między regresorami

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon.$$

Podstawiając $x_3 = x_1 x_2$ otrzymamy model liniowy postaci

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon,$$

gdzie $\beta_3 = \beta_{12}$.

ESTYMACJA PARAMETRÓW MODELU

Metoda najmniejszych kwadratów może być użyta do estymacji nieznanych parametrów modelu regresji liniowej również w przypadku wielowymiarowym.

Zakładamy, że liczba obserwacji $n > k$.

Niech y_i oznacza obserwowaną odpowiedź przy wartościach k -regresorów $x_{i1}, x_{i2}, \dots, x_{ik}$.

Dane		Regresory			
Obserwacja	Odpowiedź				
i	y	x_1	x_2	$\dots$	x_k
1	y_1	x_{11}	x_{12}	$\dots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\dots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	x_{n1}	x_{n2}	$\dots$	x_{nk}

Zakładamy, że błąd ε w modelu jest zmienną losową, dla której $E\varepsilon = 0$, $\text{Var } \varepsilon = \sigma^2$ oraz błędy dla obserwacji są nieskorelowane

UWAGA. Gdy zmienne objaśniane są niezależnymi losowymi zmiennymi, że obserwacje dla poszczególnych regresorów są niezależne oraz, że ich rozkład nie zależy od parametrów regresji (β) i σ^2 . W przypadku testowania hipotez dotyczących modelu regresji będziemy zakładać, że rozkład warunkowy y pod warunkiem $x_1, x_2, \dots, x_k$ jest rozkładem normalnym ze średnią $\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$ i wariancją σ^2 .

Model

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad i=1, 2, \dots, n.$$

Minimalizujemy funkcję

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2$$

względem $(\beta_0, \beta_1, \dots, \beta_k)$.

Estymator najmniejszych kwadratów musi spełniać

$$\frac{\partial S}{\partial \beta_0} \bigg|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) = 0$$

oraz

$$\frac{\partial S}{\partial \beta_j} \bigg|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) x_{ij} = 0$$

dla $j=1, 2, \dots, k$

W sposób bardziej przejrzysty możemy zapisać model regresji liniowej w zapisie macierzowym

$$y = X\beta + \varepsilon$$

Czyli

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Chcemy wyznaczyć wektor $\hat{\beta}$ minimalizujący

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (y - X\beta)^T (y - X\beta)$$

Zauważmy, że $S(\beta)$ możemy zapisać w postaci

$$\begin{aligned} S(\beta) &= y^T y - \beta^T X^T y - y^T X \beta + \beta^T X^T X \beta \\ &= y^T y - 2\beta^T X^T y + \beta^T X^T X \beta \end{aligned}$$

UWAGA

$\beta^T X^T y$ jest macierzą 1×1 a więc skalarzem i stąd $(\beta^T X^T y)^T = y^T X \beta$ jest tym samym skalarzem

Estymator NK musi spełniać

$$\frac{\partial S}{\partial \beta} \Big|_{\hat{\beta}} = -2X^T y + 2X^T X \hat{\beta} = 0$$

co jest równoważne warunkowi

$$X^T X \hat{\beta} = X^T y$$

Mnożąc obie strony przez macierz odwrotną $(X^T X)^{-1}$ otrzymujemy postać estymatora najmniejszych kwadratów parametru β

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

O ile macierz odwrotna dla macierzy $X^T X$ istnieje.

Dopasowany model, najmniejszych kwadratów model ma postać

$$\hat{y} = X^T \hat{\beta} = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_j,$$

$$X^T = (1, x_1, x_2, \dots, x_k).$$

Wektor dopasowanych wartości $\hat{y}_i$ odpowiadających obserwowanych wartości y_i jest równy

$$\hat{y} = X \hat{\beta} = X (X^T X)^{-1} X^T y = H y$$

gdzie macierz wymiaru $n \times n$

$$H = X (X^T X)^{-1} X^T$$

Różnica między wartościami obserwowanymi y_i i dopasowanymi $\hat{y}_i$ to reszdu (wartości resztowe) $e_i = y_i - \hat{y}_i$.

Możemy używając zapisu macierzowego napisać, że wektor wartości resztowych

$$e = y - \hat{y}$$

lub co jest równoważne

$$e = y - X\hat{\beta} = y - HY = (I - H)y$$

WŁASNOŚCI ESTYMATORA NK.

1. Estymator $\hat{\beta}$ jest estymatorem nieobciążonym

$$\begin{aligned} E(\hat{\beta}) &= E[(X^T X)^{-1} X^T y] = E[(X^T X)^{-1} X^T (X\beta + \varepsilon)] \\ &= E[(X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon] = \beta \end{aligned}$$

Ponieważ $E(\varepsilon) = 0$ oraz $(X^T X)^{-1} (X^T X) = I$

2. Wariancja a raczej macierz kowariancji wektora $\hat{\beta}$

$$\text{Cov}(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]^T$$

Ponieważ $(X^T X)^{-1} X^T$ jest macierzą stałą i wariancja y to $\sigma^2 I$ więc

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}[(X^T X)^{-1} X^T y] = (X^T X)^{-1} \text{Var}(y) [(X^T X)^{-1} X^T]^T \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1} \end{aligned}$$

ESTYMACJA σ^2

Podobnie jak w przypadku jednowymiarowym estymując σ^2 możemy opierać się na sumie kwadratów reszt

$$SS_{\text{res}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = \mathbf{e}^T \mathbf{e}$$

Podstawiając $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ otrzymujemy

$$\begin{aligned} SS_{\text{res}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} - \mathbf{y}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} \\ &= \mathbf{y}^T \mathbf{y} - 2 \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} \end{aligned}$$

Ponieważ $\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y}$ otrzymujemy

$$SS_{\text{res}} = \mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y}$$

Mozna pokazać, że nieobciążonym estymatorem σ^2 jest

$$\hat{\sigma}^2 = \frac{SS_{\text{res}}}{n - (k+1)}$$

Linia obserwacji minus linia estymowanych parametrów modelu to linia błędów swobodnych.

UWAGA. Gdy $\boldsymbol{\varepsilon} \stackrel{d}{=} N(\mathbf{0}, \sigma^2 \mathbf{I})$ to estymatorem największej wiarygodności σ^2 jest

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}$$

ESTYMATOR NAJWIĘKSZEJ WIAROGODNOŚCI

Jesli dla modelu

$$y = X\beta + \epsilon,$$

zatożymy, że błąd ϵ ma rozkład $N(0, \sigma^2 I)$ czyli $\epsilon_i, i=1, \dots, n$ są niezależnymi m. losowymi o rozkładach normalnych z gęstością

$$f(\epsilon_i) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \epsilon_i^2\right)$$

Funkcja wiarygodności ma postać

$$L(\epsilon, \beta, \sigma^2) = \prod_{i=1}^n f(\epsilon_i) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \epsilon^T \epsilon\right)$$

Ponieważ $\epsilon = y - X\beta$, funkcja wiarygodności ma więc postać

$$L(y, X, \beta, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left(-\frac{1}{2\sigma^2} (y - X\beta)^T (y - X\beta)\right)$$

Po zlogarytmowaniu

$$\ln L(y, X, \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} (y - X\beta)^T (y - X\beta)$$

Dla ustalonej wartości σ logarytm funkcji wiarygodności jest największy gdy

$$(y - X\beta)^T (y - X\beta) \quad (*)$$

jest najmniejsze. W przypadku błędów o rozkładzie normalnym estymator NW pokrywa się z estymatorem NK i ma postać

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n}$$

gdzie $\hat{\beta} = (X^T X)^{-1} X^T y$ minimizuje (*),