Algorytmiczna Analiza Danych
Ćwiczenia 4

2025-10-24

Adrian Herda
Politechnika Wrocławska

1. Zadanie 13 - Prawdopodobieństwo prawdziwej wartości w przedziale

Jako że $e_i \sim \mathcal{N}(0, \sigma^2)$ oraz są niezależne, to zgodnie z własnościami rozkładu normalnego, estymator $\hat{\beta}_1$ również ma rozkład normalny z wartością oczekiwaną równą β_1 oraz wariancją $\mathrm{Var}\left(\hat{\beta}_1\right)$.

1.1. Reguła trzech sigm

Reguła trzech sigm mówi, że w rozkładzie normalnym ok. 99.7% masy leży w przedziale $\mathbb{E}[X] \pm 3\sigma_X$ dla każdej zmiennej losowej $X \sim \mathcal{N}(\mathbb{E}[X], \sigma_X)$. Dla $\mathbb{E}[X] \pm 2\sigma_X$ mamy ok. 95%.

To tylko krótkie heurystyczne odwołanie do „regułki $68 - 95 - 99.7$” dla rozkładu normalnego ($1\sigma \approx 68\%,\ 2\sigma \approx 95\%,\ 3\sigma \approx 99.7\%$).

1.2. Ustandaryzowana zmienna

Niech

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\mathrm{Var}\left(\hat{\beta}_1\right)}} \sim \mathcal{N}(0, 1) \tag{1}$$

Wtedy:

$$\Pr\left(|\hat{\beta}_1 - \beta_1| \leq 2\sqrt{\mathrm{Var}\left(\hat{\beta}_1\right)}\right) = \Pr(|Z| \leq 2) \tag{2}$$

Ze wzoru na gęstość prawdopodobieństwa w standardowym rozkładzie normalnym mamy:

$$\begin{aligned} \Pr(|Z| \leq 2) &= \int_{-2}^{2} \varphi(x)dx \\ &= \Phi(2) - \Phi(-2) \\ &= \Phi(2) - (1 - \Phi(2)) \\ &= 2\Phi(2) - 1 \\ &= 0.9545 \end{aligned} \tag{3}$$

2. Zadanie 14 - Wzór na estymatory parametrów w modelu regresji liniowej z $k+1$ parametrami

2.1. Teza

Estymatory:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} \in \mathbb{R}^{k+1} \tag{4}$$

Wzór:

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T \vec{y} \tag{5}$$

gdzie X to macierz danych a $\vec{y}$ to wektor odpowiedzi.

2.2. Dowód

$$y = f(x) + \varepsilon \tag{6}$$

$$f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \tag{7}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k \tag{8}$$

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ X = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{9}$$

$$\begin{gathered} \vec{y} = X\beta + \varepsilon \\ \Downarrow \\ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_{11} + ... + \beta_k x_{1k} + \varepsilon_1 \\ \beta_0 + \beta_1 x_{21} + ... + \beta_k x_{2k} + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + ... + \beta_k x_{nk} + \varepsilon_n \end{pmatrix} \end{gathered} \tag{10}$$

$$\begin{gathered} \varepsilon = \vec{y} - \hat{y} \land \mathrm{RSS} = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon \\ \Downarrow \end{gathered} \tag{11}$$

$$\begin{aligned} \mathrm{RSS} &= (\vec{y} - \hat{y})^T (\vec{y} - \hat{y}) \\ J\left(\hat{\beta}\right) &= \left(\vec{y} - X\hat{\beta}\right)^T \left(\vec{y} - X\hat{\beta}\right) \\ &= \left(\vec{y}^T - \hat{\beta}^T X^T\right)\left(\vec{y} - X\hat{\beta}\right) \\ &= \vec{y}^T \vec{y} - \vec{y}^T X\hat{\beta} - \hat{\beta}^T X^T \vec{y} + X^T \hat{\beta}^T X\hat{\beta} \\ &= \vec{y}^T \vec{y} - 2\vec{y}^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta} \end{aligned} \tag{12}$$

Chcemy najmniejszą wartość RSS. Liczymy pochodną cząstkową po $\hat{\beta}$ i przyrównujemy do zera:

$$\begin{aligned} \frac{\partial J}{\partial \hat{\beta}} &= 2X^T X\hat{\beta} - 2X^T\vec{y} = 0 \\ X^T X\hat{\beta} &= X^T\vec{y} \\ \hat{\beta} &= \left(X^T X\right)^{-1} X^T\vec{y} \end{aligned} \tag{13}$$