Laboratorium nr 7 - Aproksymacja

Rozpatrywać będziemy zbiór punktów $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ (są to np. dane eksperymentalne). Niech będzie dana funkcja aproksymacyjna (dopasowująca) $F(x, p_0, p_1, \ldots, p_k)$ z nieznanymi parametrami $p_0, p_1, \ldots, p_k$. Za pomocą metody najmniejszych kwadratów dobieramy współczynniki funkcji F w taki sposób, aby

$$\sum_{i=1}^{n} \left[F(x_i, p_0, p_1, \ldots, p_k) - y_i\right]^2 = \min \tag{1}$$

1 Aproksymacja liniowa funkcji jednej zmiennej

Rozważana jest funkcja liniowa

$$y = p_0 + p_1 x. \tag{2}$$

Kryterium najmniejszych kwadratów przyjmuje w tym przypadku postać

$$S(p_0, p_1) = \sum_{i=1}^{n} (p_0 + p_1 x_i - y_i)^2 = \min \tag{3}$$

Aby wyznaczyć wartości współczynników p_0, p_1 należy rozwiązać układ równań

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix} \tag{4}$$

2 Aproksymacja funkcji jednej zmiennej za pomocą funkcji kwadratowej

Rozważana jest funkcja kwadratowa

$$y = p_0 + p_1 x + p_2 x^2. \tag{5}$$

Kryterium najmniejszych kwadratów przyjmuje w tym przypadku postać

$$S(p_0, p_1, p_2) = \sum_{i=1}^{n} \left(p_0 + p_1 x_i + p_2 x_i^2 - y_i\right)^2 = \min \tag{6}$$

Aby wyznaczyć wartości współczynników p_0, p_1, p_2 należy rozwiązać układ równań

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \end{bmatrix} \tag{7}$$

3 Ocena jakości aproksymacji

3.1 Wariancja resztowa (wariancja składnika resztowego)

$$s^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - F(x_i))^2 = \frac{1}{n-p}\sum_{i=1}^{n}\varepsilon_i^2 \tag{8}$$

y_i – wartość eksperymentalna (rzeczywista, zaobserwowana) w punkcie x_i
$F(x_i)$ – wartość przewidywana przez funkcję aproksymacyjną (wartość teoretyczna) w punkcie x_i
ε_i^2 - reszta (błąd aproksymacji) dla punktu x_i
n – liczba punktów (danych eksperymentalnych)
p – liczba dopasowanych parametrów funkcji aproksymacyjnej (dla funkcji liniowej $p = 2$, dla funkcji kwadratowej $p = 3$)

3.2 Odchylenie standardowe reszt (odchylenie standardowe składnika resztowego)

$$s = \sqrt{\frac{1}{n-p}\sum_{i=1}^{n}(y_i - F(x_i))^2} = \sqrt{\frac{1}{n-p}\sum_{i=1}^{n}\varepsilon_i^2} \tag{9}$$

3.3 MSE (mean squared error, błąd średniokwadratowy, średni błąd kwadratowy)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - F(x_i))^2 \tag{10}$$

3.4 RMSE (root mean square error)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - F(x_i))^2} \tag{11}$$

3.5 MAE (mean absolute error, błąd bezwzględny)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - F(x_i)| \tag{12}$$

3.6 Współczynnik determinacji R^2 (dla aproksymacji liniowej)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - F(x_i))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{13}$$

gdzie $\bar{y}$ to średnia z wartości obserwowanych y_i

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{14}$$