

Mały Projekt nr 4.

Temat: *Metoda najmniejszych kwadratów*

Jednym z ważnych problemów w naukach inżynierskich jest ustalenie trendu danych, czyli analiza zależności pomiędzy danymi, które pochodzą z pomiarów lub obserwacji.

Mając do dyspozycji zestaw danych w postaci zbioru n punktów $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ można próbować znaleźć taką prostą L , aby wymienione punkty leżały możliwie najbliżej tej prostej. Dobór prostej nazywany jest *regresją liniową*, a sama prosta *linią trendu*.

Popularną metodą stosowaną do rozwiązania tego problemu jest *metoda najmniejszych kwadratów*, która minimalizuje sumę kwadratów odległości pomiędzy danymi punktami a szukaną prostą.

Gdyby zadane punkty leżały na prostej $L(x) = b_0 + b_1x$, to spełniony byłby następujący układ równań liniowych:

$$(*) \quad \begin{aligned} y_1 &= b_0 + b_1x_1 \\ y_2 &= b_0 + b_1x_2 \\ &\dots \\ y_n &= b_0 + b_1x_n \end{aligned} \Leftrightarrow Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = XB$$

Najczęściej taki układ jest sprzeczny. Stąd, aby dobrać parametry b_0 oraz b_1 należy układ $(*)$ nieco zmodyfikować, mnożąc go lewostronnie przez macierz X^T . W efekcie uzyskujemy nowy układ:

$$(**) \quad X^T X B = X^T Y,$$

gdzie $X^T X \in M_n^{\mathbb{R}}(R)$.

Jeżeli macierz $X^T X$ jest odwracalna, to równanie $(**)$ ma rozwiązanie:

$$B = (X^T X)^{-1} X^T Y.$$

Otrzymane rozwiązanie minimalizuje sumę kwadratów błędów jakie pojawiają się przy rozwiązywaniu każdego z równań osobno.

Metoda najmniejszych kwadratów najczęściej jest stosowana przy regresji liniowej, ale może też być stosowana do statystycznego wyznaczania parametrów nieliniowych linii trendu. Często prostą zastępuje się wykresem wielomianu $b_0 + b_1x + b_2x^2 + \dots + b_kx^k$ stopnia k .

Zadania do wykonania:

Stosując metodę najmniejszych kwadratów:

- (a) Ustalić linię trendu dla zestawu danych pomiarowych:

$$\{(0, 4.43), (3, 6.43), (6, 8.71), (9, 9.08), (12, 11.7)\}.$$

- (b) Ustalić linię trendu dla zestawu danych pomiarowych:

$$\{(1, 18), (2, 16), (3, 13), (4, 11), (5, 9), (6, 7), (7, 5), (8, 4), (9, 1), (10, 1)\}.$$

- (c) Znaleźć funkcję kwadratową oddającą trend dla podanego zestawu danych:

$$\{(1, 7), (2, 10), (3, 11), (4, 24), (5, 34), (6, 46), (7, 55)\}.$$

- (d) Dobrać prostą a następnie wielomian stopnia $k = 3$ najlepiej przybliżające zestaw danych pomiarowych:

$$\{(1, 2.54968), (2, 2.57332), (3, 3.77028), (4, 4.50018), (5, 6.23465)\}.$$