0.1 Prosta regresji.

Załóżmy, że w wyniku pomiarów wielkości $y(x)$ zależnej od argumentu x w punktach

$$x_0,\ x_1\ x_2,\ ...,\ x_m$$

otrzymaliśmy wartości

$$y_0,\ y_1\ y_2,\ ...,\ y_m,$$

które zanotowaliśmy w postaci tabeli

Tabela danych punktów (x, y)

x	x_0	x_1	x_2	....	x_m
y	y_0	y_1	y_2	....	y_m

Na przykład zmierzyliśmy temperetarę w każdą poniedziałek września, 2018 i zanotowaliśmy wyniki pomiaru w tabeli

Tabela 1

dzień	1	8	15	22	29
temperatuta	15	17	20	14	10

W tym doświadczeniu $m = 4$

(1)

0.2 Prosta regresji

Prosta regresji o równaniu

$$y(x) = a_1 x + a_0$$

jest najlepiej dopasowaną prostą do danych punktów $(x_i, y_i),\ i = 0, 1, 2, ..., m$ w sensie *najmniejszych kwadratów.*
Prosta jest najlepiej dopasowaną do danych punktów w tabeli w *sensie najmniejszych kwadratów*, jeżeli suma kwadratów odległości wszystkich punktów od prostej położonej na płaszczyźnie kartezjańskiej jest najmniesza.
Współczynniki $a_0,\ a_1$ obliczmy rozwiązując następujący układ dwóch równań

$$s_0 a_0 + s_1 a_1 = v_0$$

$$s_1 a_0 + s_2 a_1 = v_1$$

gdzie

$$
\begin{aligned}
s_0 &= m+1,\\
s_1 &= x_0 + x_1 + x_2 + \cdots + x_m = \sum_{i=0}^{m} x_i,,\\
s_2 &= x_0^2 + x_1^2 + x_2^2 + \cdots + x_m^2 = \sum_{i=0}^{m} x_i^2,\\
v_0 &= y_0 + y_1 + y_2 + \cdots + y_m = \sum_{i=0}^{m} y_i,\\
v_1 &= x_0 * y_0 + x_1 * y_1 + x_2 * y_2 + \cdots + x_m * y_m = \sum_{i=0}^{m} x_i * y_i
\end{aligned}
$$

Układ dwóch równań o współczynnikach s_0, s_1, s_2 i prawych stronach v_0, v_1 wynika z warunku na minimum funkcji

$$
\begin{aligned}
\phi(a_0, a_1) &= (a_0 + a_1 * x_0 - y_0)^2 + (a_0 + a_1 * x_1 - y_1)^2 + \cdots + (a_0 + a_1 * x_m)^2\\
&= \sum_{i=0}^{m} (a_0 + a_1 x_i - y_i)^2,
\end{aligned}
$$

dwóch zmiennych a_0, a_1.
Warunek na minimum funkcji dwóch zmiennych jest sformułowany w pojciach pochodnych cząstkowych funkcji wielu zmiennych i wykracza poza zakres programu szkół średnich. Sam układ dwóch równań jest prosty i łatwo rozwiązywany metodą eliminacji niewiadomych a_0, a_1.

Przykład 0.1 *Znajdź równanie prostej*

$$
y(x) = a_1 * x + a_0
$$

najlepiej dopasowane do danych w tabeli (1) i podaj jej wykres we współrzędnych kartezjańskich x, y.
Oblicz temperaturę 20-go września, 2018 roku $y(20)$ z równania prostej regresji.

Rozwiązanie.
Obliczamy współczynniki s_0, s_1, s_2 i prawe strony $v_0,\ v_1$ dla $m = 4$ i danych w tabeli (1)

$$
\begin{aligned}
&s_0 = m + 1 = 4 + 1 = 5,\\
&s_1 = x_0 + x_1 + x_2 + \cdots + x_m = 1 + 8 + 15 + 22 + 29 = 75,\\
&s_2 = x_0^2 + x_1^2 + x_2^2 + \cdots + x_m^2 = 1 * 15 + 8 * 17 + 15 * 20 + 22 * 14 + 29 * 10 = 1615,\\
&v_0 = y_0 + y_1 + y_2 + \cdots + y_m = 15 + 17 + 20 + 14 + 10 = 76,\\
&v_1 = x_0 * y_0 + \cdots + x_m * y_m = 1 * 15 + 8 * 17 + 15 * 20 + 22 * 14 + 29 * 10 = 1049
\end{aligned}
$$

Dla danych w tabeli (1) otrzymamy następujący układ dwóch równań

$$5a_0 + 75a_1 = 76$$

$$75a_0 + 1615a_1 = 1049$$

Rozwiązanie układu metodą eliminacji Gaussa.
Pomnóżmy obie strony pierwszego równanie przez 15 i odejmijmy stronami równanie drugie od równania pierwszego, jak niżej

$$\begin{array}{ll} 5a_0 + 75a_1 = 76 \ \ | * 15, & 75a_0 + 1125a_1 = 1140, \\ 75a_0 + 1615a_1 = 1049, & 75a_0 + 1615a_1 = 1049, \\ -------------- & ----------- \\ 1125a_1 - 1615a_1 = 1140 - 1049, & -490a_1 = 91 \end{array}$$

Z równania $-490a_1 = 91$ obliczamy

$$a_1 = \frac{91}{-490} = -\frac{13}{70} = -0.185714$$

z równania $5a_0 + 75a_1 = 76$ obliczamy

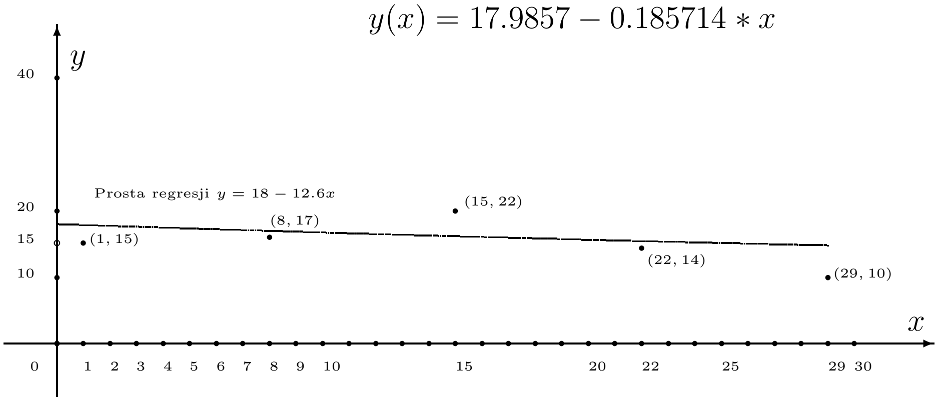
$$a_0 = \frac{76 + 75 * \frac{13}{70}}{5} = \frac{1259}{70} = 17.9857$$

Równanie prostej regresji najlepiej dopasowanej do danych w tabeli (1) w sensie metody najmniejszych kwadratów

$$y = \frac{1259}{70} - \frac{13}{70}x, \qquad lub \qquad y = 17.9857 - 0.185714x$$

Temperatura 20-go września $y(20) = 17.9857 - 0.185714 * 20 = 14.18856$.
Wykres prostej regresji

$$y(x) = 17.9857 - 0.185714 * x$$



Zauważmy, że żaden z danych punktów w tabeli (1) nie leży na prostej regresji. Pomimo tego prosta jest najlepiej dopasowana do tych danych w sensie metody najmniejszych kwadratów.

0.3 Zadania

Zadanie 0.1 *Wzrost chłopca zanowtowano w tableli*

Tabela 2

rok	1 *rok*	5 *lat*	10 *lat*	15 *lat*	20 *lat*
wzrost	80*cm*	100*cm*	140*cm*	160*cm*	180*cm*

W tym doświadczeniu $m = 4$

- *Znajdź równanie*

$$y(x) = a_1 x + a_0$$

prostej regresji dla danych w powyższej tableli 2.

- *Na płaszczyźnie katezjańskiej we współrzdnych* x, y *zaznacz punkty*

$$(x_0, y_0) = (1, 80), \qquad (x_1, y_1) = (5, 100),$$

$$(x_2, y_2) = (10, 140), \quad (x_3, y_3) = (15, 160),$$

$$(x_4, y_4) = (20, 180)$$

i podaj wykres prostej regresji.

- *Oblicz wzrost chłopca* $y(18)$ *w wieku 18 lat z prostej regresji.*

Prof. dr Tadeusz STYŠ Warszawa, 18 wrzesień 2018.