

Linearna regresija – hipoteza je

$$h_w(x) = w_1x + w_0.$$

Tražimo vrijednost minimuma za sljedeću funkciju:

$$Loss(h_w) = \sum_{j=1}^N L_2(y_j, h_w(x_j)) = \sum_{j=1}^N (y_j - h_w(x_j))^2 = \sum_{j=1}^N (y_j - (w_1x_j + w_0))^2$$

Diferenciranjem gornjeg izraza po nepoznatim w_1 i w_0 , dobijamo dvije jednačine sa dvije nepoznate.

$$\frac{\partial}{\partial w_0} \sum_{j=1}^N (y_j - (w_1x_j + w_0))^2 = 0$$

$$\frac{\partial}{\partial w_1} \sum_{j=1}^N (y_j - (w_1x_j + w_0))^2 = 0$$

Lako se pokazuje da ovaj sistem ima jedinstveno rješenje (npr. primjenom Kramerovog pravila ili Gausovom eliminacijom). Rješenja sistema mogu se predstaviti formulama:

$$w_1 = \frac{N(\sum x_j y_j) - (\sum x_j)(\sum y_j)}{N(\sum x_j^2) - (\sum x_j)^2}; \quad w_0 = (\sum y_j - w_1(\sum x_j))/N$$

Primjer: Dati su sljedeći ulazni podaci:

X	Y
60	3.1
61	3.6
62	3.8
63	4
65	4.1

Izračunamo tražene veličine:

X	Y	X*Y	X*X
60	3.1	60 * 3.1 = 186	60 * 60 = 3600
61	3.6	61 * 3.6 = 219.6	61 * 61 = 3721
62	3.8	62 * 3.8 = 235.6	62 * 62 = 3844
63	4	63 * 4 = 252	63 * 63 = 3969
65	4.1	65 * 4.1 = 266.5	65 * 65 = 4225

Sume su: $\sum x_j = 311$, $\sum y_j = 18.6$, $\sum x_j y_j = 1159.7$, $(\sum x_j)^2 = 19359$. Uvrstimo u formulu i dobijamo:

$$w_1 = 0.19, \quad w_0 = -8.098 \text{ tj.}$$

$$y = -8.098 + 0.19x.$$