

Jasne — rozbijmy to **na elementy** i pokażmy **dokładnie algebraicznie** i **intuicyjnie**, dlaczego przy cofaniu standaryzacji współczynników dzielimy przez  $\sigma$  (a bias trzeba skorygować). Zostanę przy prostym przykładzie z jednym featurem, potem uogólnimy na wiele cech.

## 1) Przypadek najprostszy — jedna cecha + bias

Mamy oryginalne dane:

- $x$  — pojedyncza cecha,
- $1$  — kolumna bias (wyraz wolny),
- $y$  — target.

Model liniowy na oryginalnych cechach:

$$\hat{y} = \theta_0 + \theta_1 x.$$

Teraz zeskalamy cechę  $x$  (standaryzacja):

$$x_{\text{scaled}} = \frac{x - \mu}{\sigma},$$

gdzie  $\mu = \mathbb{E}[x]$ ,  $\sigma = \text{std}(x)$ .

Model trenowany na zeskaliwanej cesze to:

$$\hat{y} = \theta_0^{(s)} + \theta_1^{(s)} x_{\text{scaled}}.$$

Podstawmy definicję  $x_{\text{scaled}}$ :

$$\hat{y} = \theta_0^{(s)} + \theta_1^{(s)} \frac{x - \mu}{\sigma}.$$

Rozbijamy sumę:

$$\hat{y} = \theta_0^{(s)} + \frac{\theta_1^{(s)}}{\sigma} x - \frac{\theta_1^{(s)} \mu}{\sigma}.$$

Teraz porównaj to z formą  $\hat{y} = \theta_0 + \theta_1 x$ . Widać, że

$$\theta_1 = \frac{\theta_1^{(s)}}{\sigma}$$

oraz że wyraz wolny (bias) musi być

$$\theta_0 = \theta_0^{(s)} - \frac{\mu}{\sigma} \theta_1^{(s)}.$$

**Intuicja jednostkowa:** jeżeli  $x$  zostało podzielone przez  $\sigma$ , to „jednostka” zmiany  $x_{\text{scaled}}$  to teraz  $\sigma$  jednostek oryginalnego  $x$ . Współczynnik  $\theta_1^{(s)}$  odpowiada zmianie  $y$  na jedną jednostkę zmiany  $x_{\text{scaled}}$  — czyli na  $\sigma$  jednostek oryginalnego  $x$ . Aby uzyskać współczynnik mierzący zmianę  $y$  na jedną jednostkę oryginalnego  $x$ , musisz **podzielić** przez  $\sigma$ .

## Podsumowując:

### 3) Intuicyjne wyjaśnienie „dlaczego dzielimy przez $\sigma$ ”

- Po standaryzacji, zmiana o **1** w  $x_{\text{scaled}}$  odpowiada zmianie o  $\sigma$  w oryginalnym  $x$ .
- $\theta_1^{(s)}$  mówi „o ile zmienia się  $y$ , gdy  $x_{\text{scaled}}$  zwiększy się o 1 (jednostkę zeskalowaną)”.
- Aby uzyskać „o ile zmienia się  $y$ , gdy oryginalne  $x$  zwiększy się o 1 (jednostkę oryginalną)”, musimy podzielić tę zmianę przez  $\sigma$ . Stąd  $\theta_1 = \theta_1^{(s)} / \sigma$ .