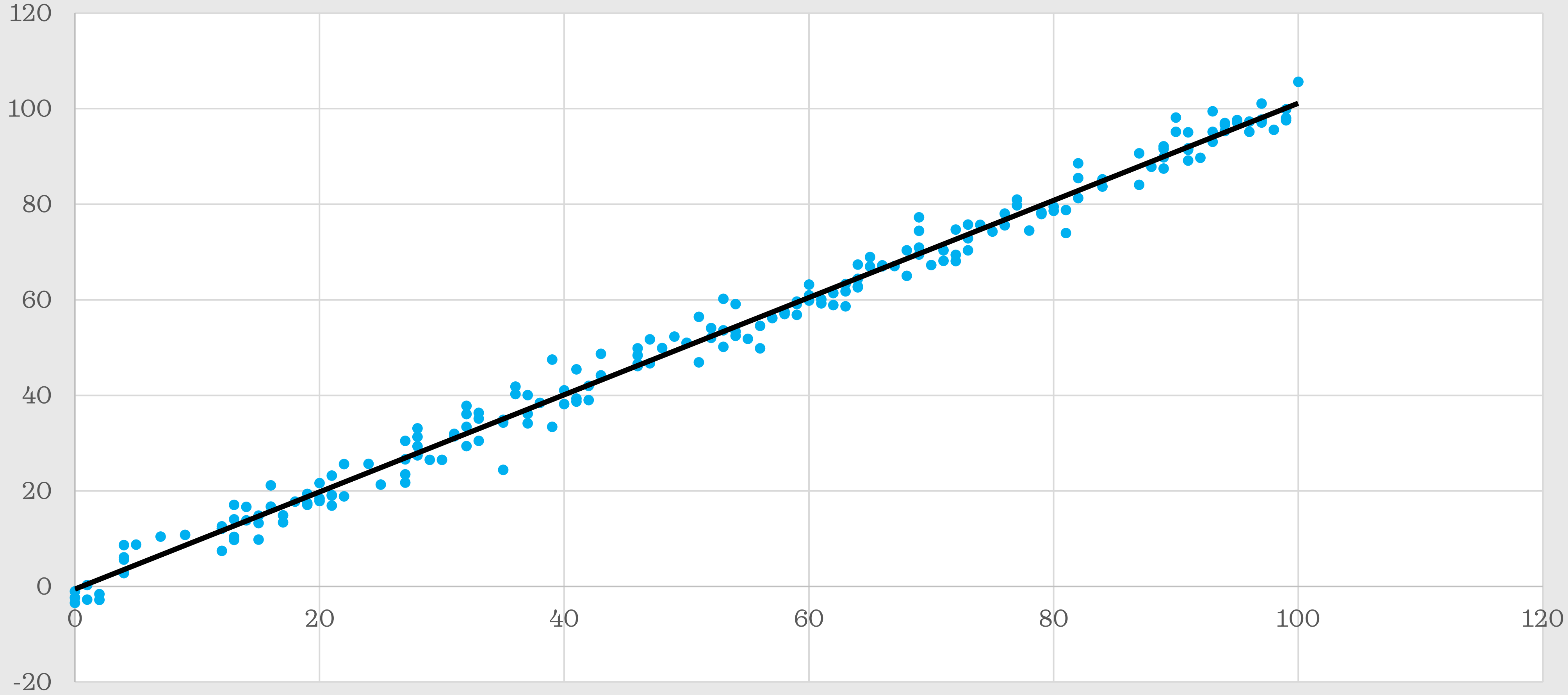
REGRESJA LINIOWA

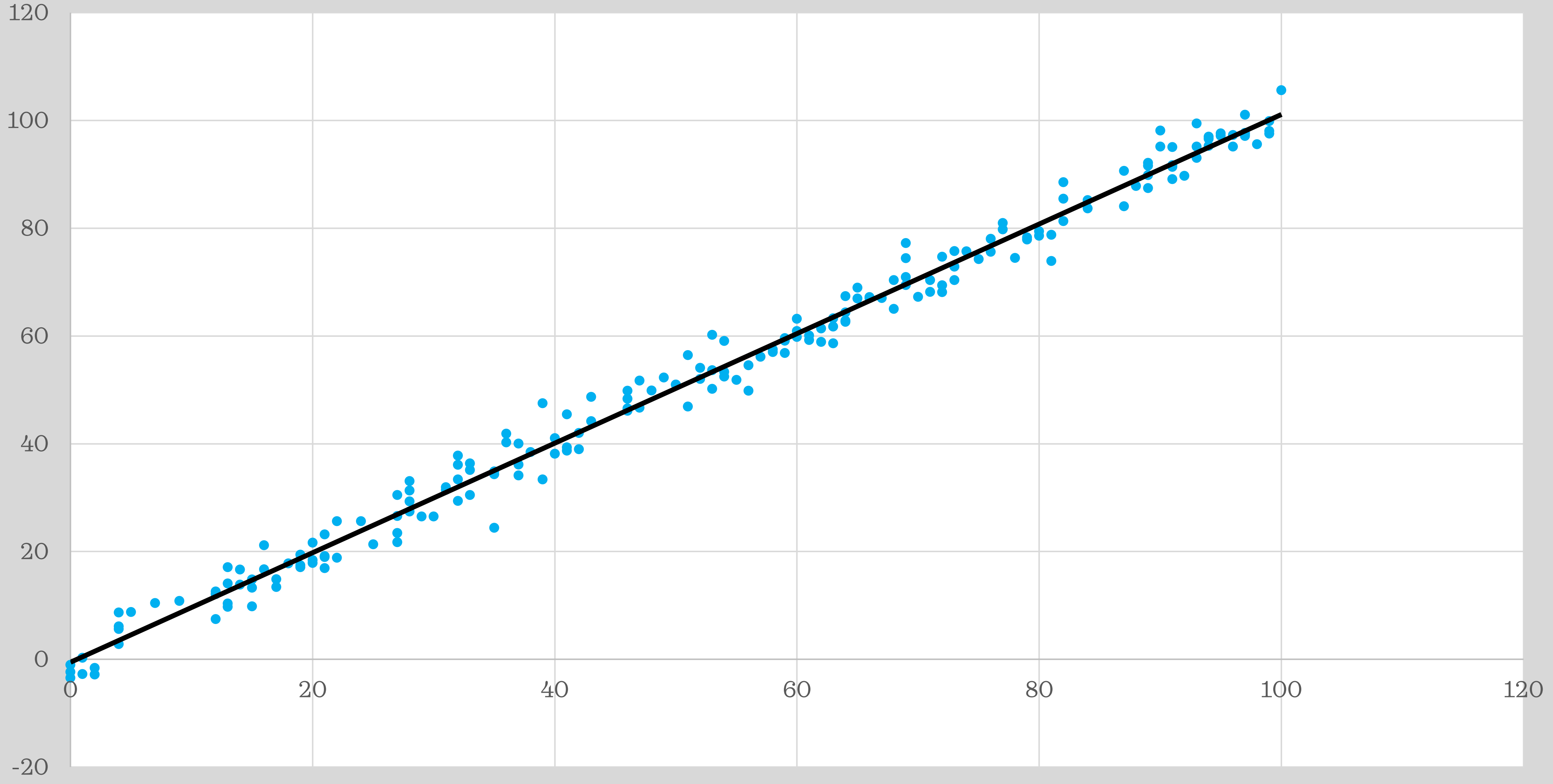
Golem bootcamp 2021

Problem regresji liniowej

Mamy zbór punktów i chcemy dopasować do niego linię, która będzie najlepiej opisywać wszystkie dane



$$y = \beta_0 + \beta_1 x$$



Mean Square Error – funkcja straty

Jak sprawdzić jak daleko jesteśmy od prawdziwego wyniku?

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - pred_i)^2$$

Zmniejszamy β_0 i β_1 o stałą wartość α

Dlaczego to nie zadziała?

- Jeśli α jest za duża to będziemy mieli duży error
- Jeśli α jest za mała to będziemy mieli długi czas liczenia

Liczymy pochodne cząstkowe z funkcji straty

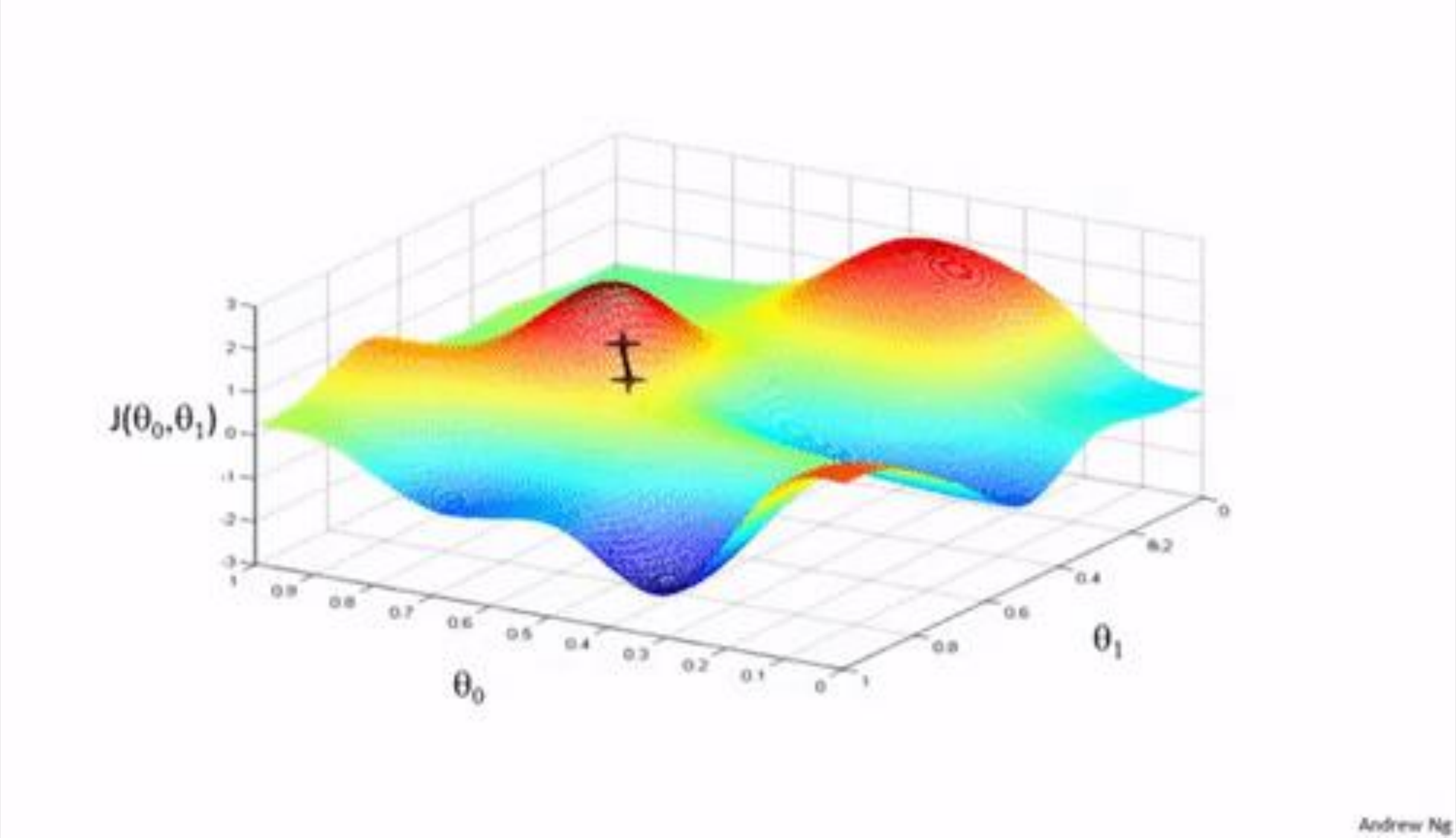
Pochodna po β_0:

$$D_0 = \frac{-2}{n}\sum_{i=0}^{n}(y_i - (\beta_0 + \beta_1 x_i))$$

Pochodna po β_1:

$$D_1 = \frac{-2}{n}\sum_{i=0}^{n}x_i(y_i - (\beta_0 + \beta_1 x_i))$$

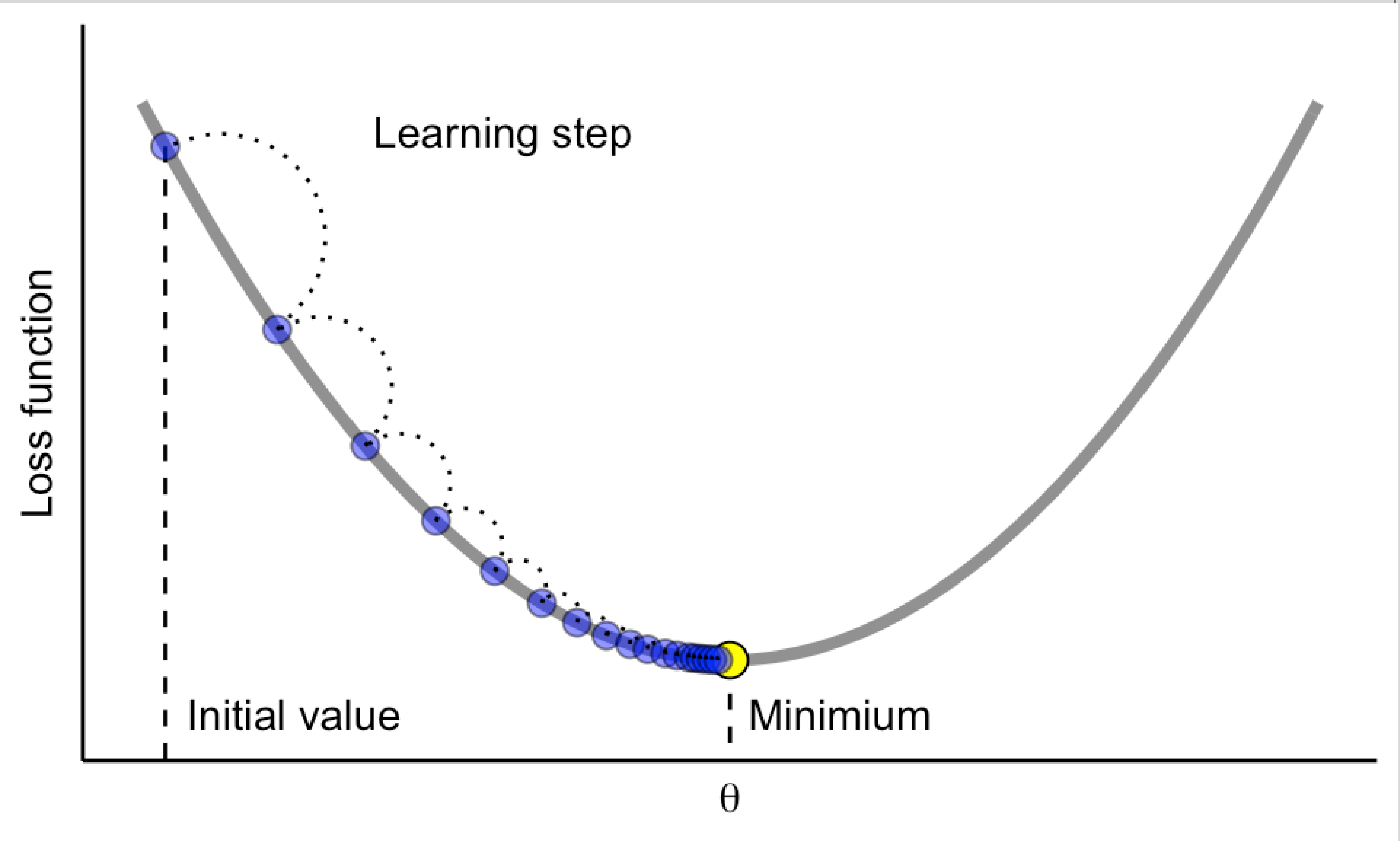
Spadek gradientu



Używamy gradientu i współczynnika α

$\beta_0 = \beta_0 \ - \alpha\, D_0$

$\beta_1 = \beta_1 \ - \alpha\, D_1$



Więcej niż jedna wartość

y = $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$

Wtedy:

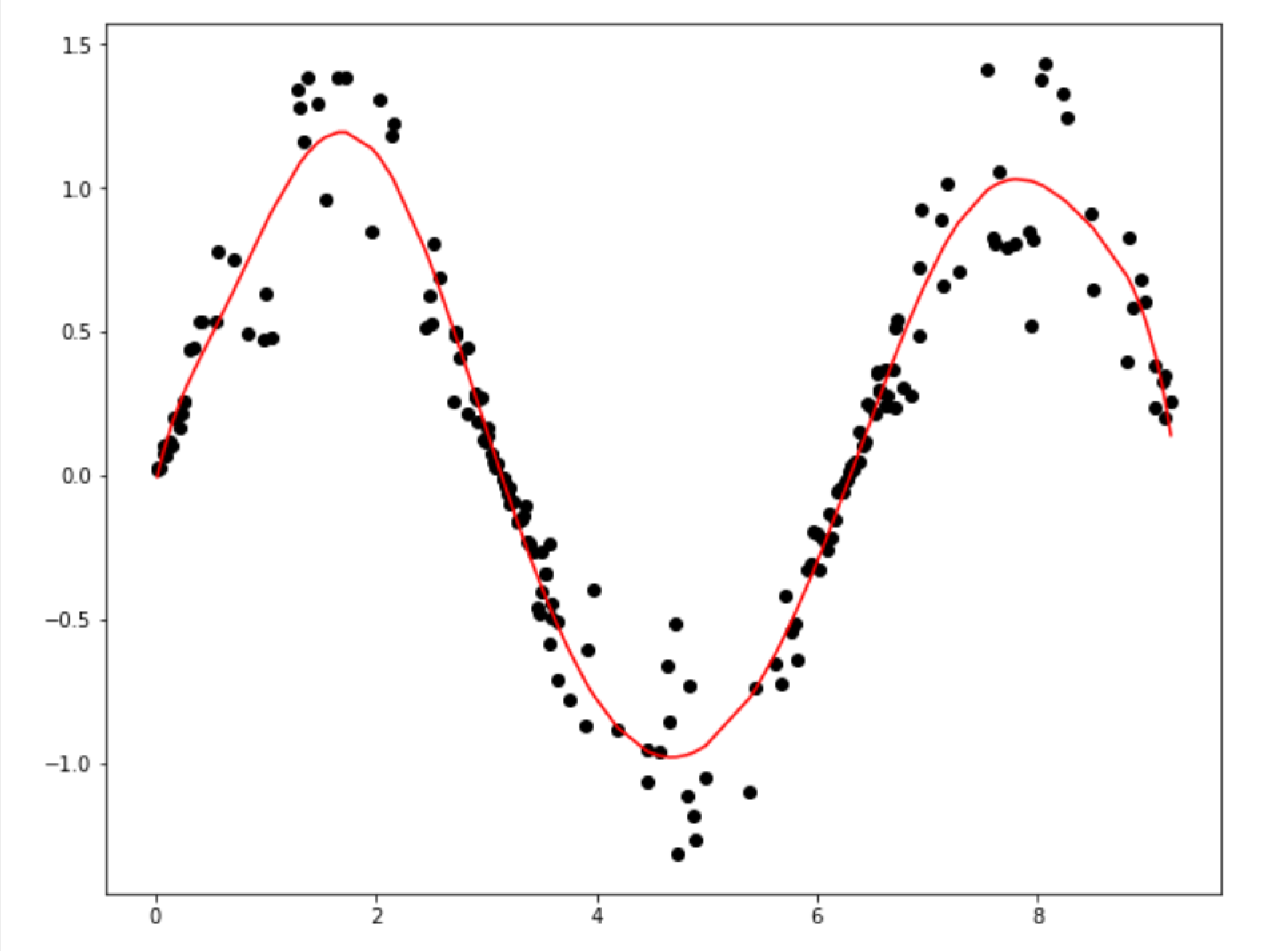
$\beta_0 = \beta_0$ - α D_0

$\beta_1 = \beta_1$ - α D_1

…

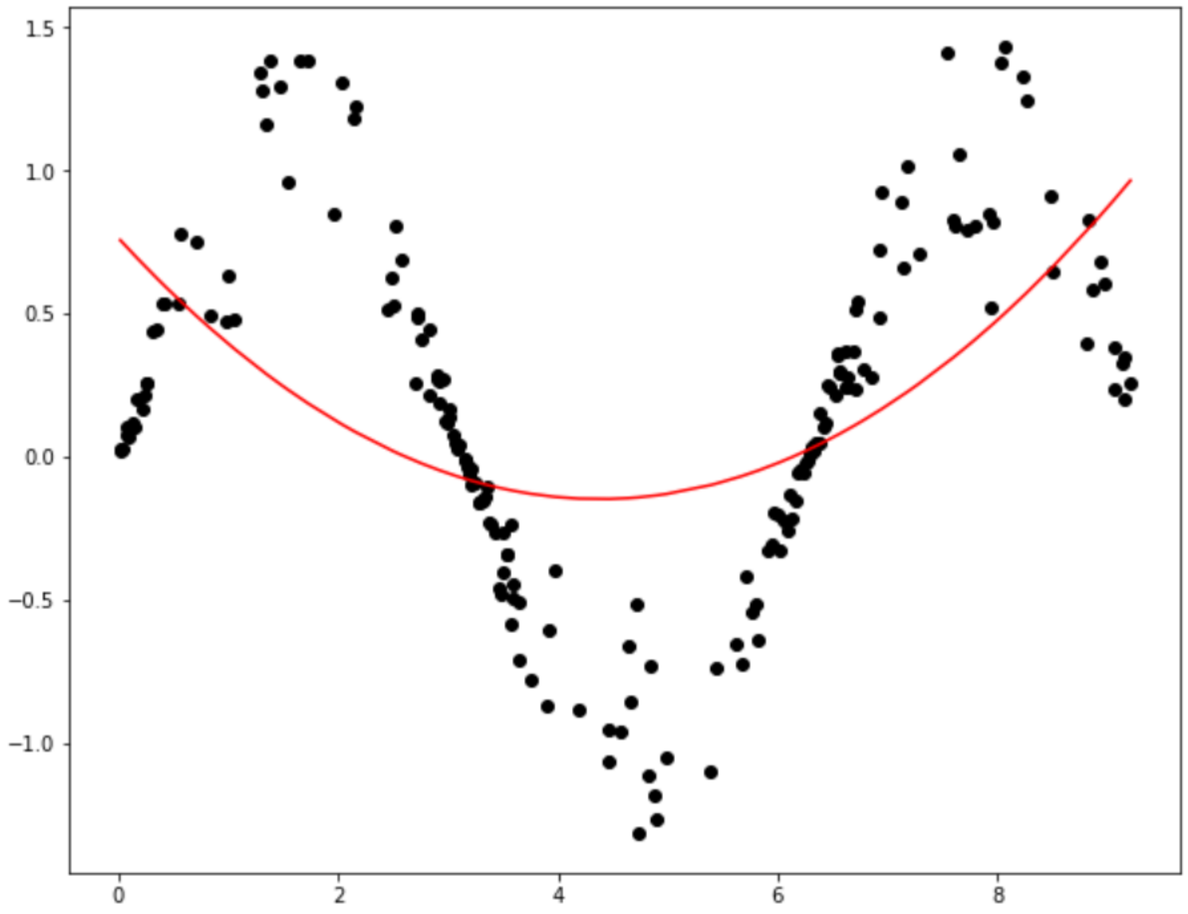
$\beta_p = \beta_p$ - α D_p

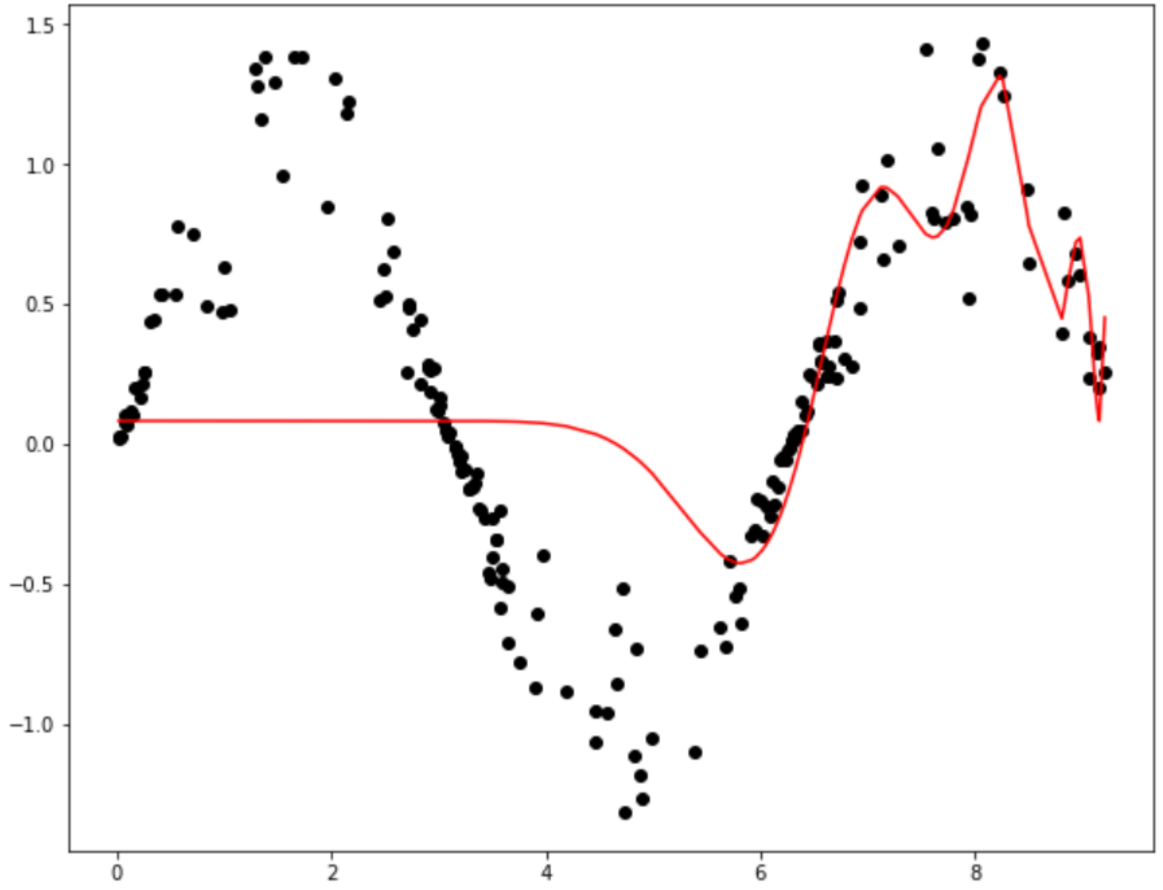
Inne rodzaje regresji – regresja wielomianowa



$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_p X^p$$

underfitting & overfitting





Regularyzacja – zapobiegamy overfittingowi

- Jeśli β_0 , β_1 ,β_2 , β_3 wynoszą kolejno 0.1, 0.2, 0.04, 500 to β_3 ma decydujący wpływ na wynik

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - pred)^2 + \lambda \sum_{j=0}^{p}{\beta_j}^2 \text{ – ridge regularization}$$

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - pred)^2 + \lambda \sum_{j=0}^{p}\left|\beta_j\right| \text{ – lasso regularization}$$