Zacznijmy od rozdzielania gradientu całej funkcji kosztu na osobne czynniki, korzystając z własności że pochodna sumy jest sumą pochodnych:

$$\begin{aligned}
\nabla_w L(X,y,w) &= \nabla_w\Big(\frac{1}{N}\sum_i (w^T x_i - y_i)^2 + \alpha \sum_i w_i^2\Big) \\
&= \frac{1}{N}\sum_i \nabla_w (w^T x_i - y_i)^2 + \nabla_w \alpha \sum_i w_i^2 \\
&= \frac{1}{N}\sum_i \nabla_w ((w^T x_i)^2 - 2w^T x_i y_i + y_i^2) + \nabla_w \alpha \sum_i w_i^2 \\
&= \frac{1}{N}\sum_i \nabla_w (w^T x_i)^2 - \nabla_w (2w^T x_i y_i) + \nabla_w (y_i^2)) + \nabla_w \alpha \sum_i w_i^2
\end{aligned} \tag{1}$$

Teraz policzmy osobno pochodne:

$$\begin{aligned}
\nabla_w (w^T x_i)^2 &= \frac{\partial (w^T x_i)^2}{\partial w^T x_i} \cdot \frac{\partial w^T x_i}{\partial w} \qquad \text{(reguła łańcuchowa)} \\
&= 2(w^T x_i) \cdot x_i \\
\nabla_w 2w^T x_i y_i &= 2y_i \nabla_w w^T x_i = 2y_i x_i \\
\nabla_w y_i^2 &= 0 \\
\nabla_w \alpha \sum_i w_i^2 &= \alpha \nabla_w w^T w = 2\alpha w
\end{aligned} \tag{2}$$

Składając w całość:

$$\begin{aligned}
\nabla_w L(X,y,w) &= \frac{1}{N}\sum_i [2(w^T x_i) \cdot x_i - 2y_i x_i] + 2\alpha w \\
&= \frac{2}{N}\sum_i [(w^T x_i - y_i) x_i] + 2\alpha w
\end{aligned} \tag{3}$$