Parameter Update

$$\theta^{t+1} = \theta^t - \lambda \frac{\partial L}{\partial \theta}, \quad \lambda = \text{learning rate.}$$

→ Full batch
→ Stochastic / single sample.
→ Mini-batch

$$\theta^{t+1} = \theta^t - \frac{\lambda}{B} \sum_{i=1}^{B} \frac{\partial L^i}{\partial \theta}$$

Gradient Checking:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}$$

input: $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$ $x^+ = \begin{bmatrix} x_1 + h_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$, $x^- = \begin{bmatrix} x_1 - h_1 \\ \vdots \\ x_N \end{bmatrix}$