Gradient Descent

Gradient descent, or back propagation, is a first order iterative optimization algorithm used to minimize the cost function when training artificial neural networks. As stated in $Recurrent Neural Networks (Mandic, Chambers)$:

Gradient based learning aims to update iteratively the weight vector w of an adaptive system so that a non-negative error measure $J(\cdot)$ is reduced at each time step k,

$$J(w + \Delta w) < J(w) \tag{1}$$

where Δw is the change in w from one iteration to the next. Using a Taylor series expansion to approximate the error measure,

$$J(w) + \Delta w \frac{\partial J(w)}{\partial w} + O(w^2) < J(w) \tag{2}$$

With the assumption that the higher-order terms in the left-hand side of (2) can be neglected, (1) can be rewritten as

$$\Delta w \frac{\partial J(w)}{\partial w} + O(w^2) < 0 \tag{3}$$

From (3), an algorithm would continuously reduce the error measure on the run, should change the weights in the opposite direction of the gradient $\partial J(w)/\partial w$,

$$\Delta w = -n \frac{\partial J(w)}{\partial w} \tag{4}$$