Mathematics and Artificial Neural Network

Giorgi Kakulashvili

February 14, 2025

1. Linear Function Approximation

$$T(x) = f\left(Wx + b\right) \tag{1.1}$$

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad T = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{pmatrix}.$$

$$t_i = f(y_i) = f\left(\sum_{k=1}^{m} w_{ik}x_k + b_i\right), \quad \frac{\partial t_i}{\partial w_{ik}} = x_k f'(y_i), \quad \frac{\partial t_i}{\partial b_i} = f'(y_i).$$

Let's define loss function L as

$$L(\hat{t}; t) = L(\hat{t}_1, \ldots, \hat{t}_m; t_1, \ldots, t_m) = \frac{1}{2}\sum_{i=1}^{m}\left(\hat{t}_i - t_i\right)^2. \tag{1.2}$$

where t is original vector and $\hat{t}$ is computed using (1.1).

$$\delta w_{ij} = \frac{\partial L}{\partial w_{ij}} = \sum_{i=1}^{m}\left(\hat{t}_i - t_i\right)\frac{\partial \hat{t}_i}{\partial w_{ij}} = \sum_{i=1}^{m}\left(f(y_i) - t_i\right)x_j f'(y_i),$$

$$\delta b_i = \frac{\partial L}{\partial b_i} = \sum_{i=1}^{m}\left(\hat{t}_i - t_i\right)\frac{\partial \hat{t}_i}{\partial b_i} = \sum_{i=1}^{m}\left(f(y_i) - t_i\right)f'(y_i).$$

Therefore, we obtain following

$$w_{ij} \leftarrow w_{ij} - \gamma\delta w_{ij}, \tag{1.3}$$

$$b_i \leftarrow b_i - \gamma\delta b_i, \tag{1.4}$$

where γ is some small positive constant.