Gradient calculation in multilayer percetrons

Tuomas Virtanen, Tampere University

October 2025

Indices: l is the layer index, L is the number of layers, i and j are indices of neurons.

1 Forward pass

Network inputs:

$$a_j^0 = x_j \tag{1}$$

Linear layer pre-activations

$$z_j^l = \sum_i w_{ij}^l a_i^{l-1} + b_j^l \tag{2}$$

Non-linearity:

$$a_j^l = f(z_j^l) \tag{3}$$

Output layer:

$$\hat{y}_j = z_j^L \tag{4}$$

Squared loss for one sample:

$$c = \sum_j (\hat{y}_j - y_j)^2 \tag{5}$$

2 Gredient calculations

$$\frac{\partial c}{\partial w_{ij}^l} = \frac{\partial c}{\partial z_j^l} \frac{\partial z_j^l}{\partial w_{ij}^l} \tag{6}$$

$$\frac{\partial c}{\partial b_j^l} = \frac{\partial c}{\partial z_j^l} \frac{\partial z_j^l}{\partial b_j^l} \tag{7}$$

where

$$\frac{\partial z_j^l}{\partial w_{ij}^l} = a_i^{l-1} \tag{8}$$

$$\frac{\partial z_j^l}{\partial b_j^l} = 1 \tag{9}$$

and

$$\frac{\partial c}{\partial z_j^l} = \frac{\partial c}{\partial a_j^l} \frac{\partial a_j^l}{\partial z_j^l} \tag{10}$$

where

$$\frac{\partial a_j^l}{\partial z_j^l} = f'(z_j^l) \tag{11}$$

and

$$\frac{\partial c}{\partial a_j^l} = \sum_i \frac{\partial c}{\partial z_i^{l+1}} \frac{\partial z_i^{l+1}}{\partial a_j^l} \tag{12}$$

where

$$\frac{\partial z_i^l}{\partial a_j^l} = w_{ji} \tag{13}$$

For the output layer:

$$\frac{\partial c}{\partial z_j^L} = \hat{y}_j - y_j \tag{14}$$

3 Backpropagation

Before the backpropagation algorithms, the forward pass is executed to compute the quantities in equations 2 - 5.

The backpropagation algorithms consists of computing first the gradient of the output layer 14 and then recursively the gradients of the previous layer outputs in Eq. 12 and pre-activations in Eq. 10 for layers from $l-1$ to 1, and finally the gradients of the weights and biases based on Eqs. 6 and 7.