

Gradient calculation in multilayer perceptrons

Tuomas Virtanen, Tampere University

October 2024

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

1 Forward pass

Network inputs:

$$z_j^0 = x_j \quad (1)$$

Linear layer pre-activations

$$a_j^l = \sum_i w_{ij}^l z_i^{l-1} + b_j^l \quad (2)$$

Non-linearity:

$$z_j^l = f(a_j^l) \quad (3)$$

Output layer:

$$\hat{y}_j = a_j^L \quad (4)$$

Squared loss for one sample:

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

2 Gradient calculations

$$\frac{\partial c}{\partial w_{ij}^l} = \frac{\partial c}{\partial a_j^l} \frac{\partial a_j^l}{\partial w_{ij}^l} \quad (6)$$

$$\frac{\partial c}{\partial b_j^l} = \frac{\partial c}{\partial a_j^l} \frac{\partial a_j^l}{\partial b_j^l} \quad (7)$$

where

$$\frac{\partial a_j^l}{\partial w_{ij}^l} = z_j^{l-1} \quad (8)$$

$$\frac{\partial a_j^l}{\partial b_j^l} = 1 \quad (9)$$

and

$$\frac{\partial c}{\partial a_j^l} = \frac{\partial c}{\partial z_j^l} \frac{\partial z_j^l}{\partial a_j^l} \quad (10)$$

where

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

and

$$\frac{\partial c}{\partial z_j^l} = \sum_i \frac{\partial c}{\partial a_i^{l+1}} \frac{\partial a_i^{l+1}}{\partial z_j^l} \quad (12)$$

where

$$\frac{\partial a_i^l}{\partial z_j^l} = w_{ji} \quad (13)$$

For the output layer:

$$\frac{\partial c}{\partial a_j^L} = \hat{y}_j - y_j \quad (14)$$

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.