

Theoretical task 9.

Recommendations: all solutions should be short, mathematically strict (unless qualitative explanation is needed), precise with respect to the stated question and clearly written. Solutions may be submitted in any readable format, including images.

Submission link: [here](#)

1. Consider a neural network with a single hidden layer and sigmoid activation functions. Explain why initializing all weight with zero values in Back Propagation algorithm is a bad idea.
2. Count the number of parameters **after each layer** in the following NN setting:
 - Input layer: 8 neurons
 - Hidden layer 1: 15 neurons with sigmoid activations
 - Hidden layer 2: 7 neurons with ReLU activations
 - Output layer: 3 neurons

Don't forget to include bias term in your calculations

3. Consider classification task with $C > 2$ classes.
 - (a) How many output neurons would corresponding NN have?
 - (b) Consider cross-entropy loss function for a single object x

$$L = - \sum_{c=1}^C y_c \log(\hat{y}_c),$$

where $y_c = 1$ if object x belongs to class c and 0, otherwise and \hat{y}_c is estimated probability for class c , calculated with softmax function:

$$\hat{y}_c = \frac{e^{I_c}}{\sum_{k=1}^C e^{I_k}}, \quad I_1, \dots, I_C \text{ are inputs of output layer}$$

Calculate δ_j for output neuron j both for $j = c$ and $j \neq c$.