Theoretical task 8.

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.

1. Consider the following sample with two classes:

x_i	x_2	y
0	0	1
-1	1	1
1	1	1
0	1	0

Table 1: Sample example.

Suppose you have a neural network with structure is shown in Fig. 1.

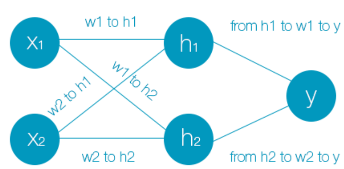

Figure 1: Neural network structure.

Find weights (w_{11}, w_{12}, w_{21}, ...), free coefficients (w_{01}, w_{02}, ...) and activation function of this network to separate two classes of the sample in Tab. 1.

2. Consider a binary classification task with sample is shown in Fig. 2. To solve this task you use a neural network which is demonstrated in Fig. 1. Suppose you use sigmoid activation function. What is the minimal number of neurons of the hidden layer needed to separate two classes in in Fig. 2? Why?

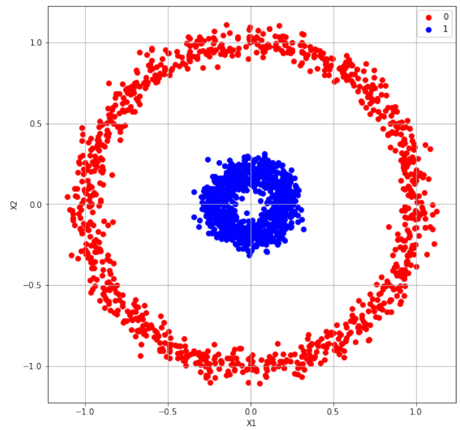

Figure 2: Two circles sample.

3. Consider binary classification task and a neural network with one output neuron for this task. The output neuron has sigmoid activation function:

$$\hat{y} = \frac{1}{1 + e^{-a}} \tag{1}$$

where $\hat{y}$ is output value of the output neuron, a is the neuron's value before sigmoid function calculation.

Consider a binary cross-entropy error function:

$$L = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}) \tag{2}$$

where $y \in \{0, 1\}$ is true value. Prove the following:

$$\frac{\partial L}{\partial a} = \hat{y} - y \tag{3}$$

4. Consider an neural network for multiclassification task with C output neurons with softmax activation function. Each output neuron corresponds to one of C classes and returns the following value:

$$\hat{y}_i = \frac{e^{a_i}}{\sum_{j=1}^{C} e^{a_j}} \tag{4}$$

where $\hat{y}_i$ is output value of the i-th output neuron, a_j is the j-th neuron value before softmax function calculation.

Consider a cross-entropy error function:

$$L = -\sum_{i=1}^{C} y_i \log \hat{y}_i \tag{5}$$

where $y_i \in \{0, 1\}$ is true value for the i-th neuron. Prove the following:

$$\frac{\partial L}{\partial a_i} = \hat{y}_i - y_i \tag{6}$$