1. Consider the neural network with 3 layers shown in Figure 1, which is used for regression.

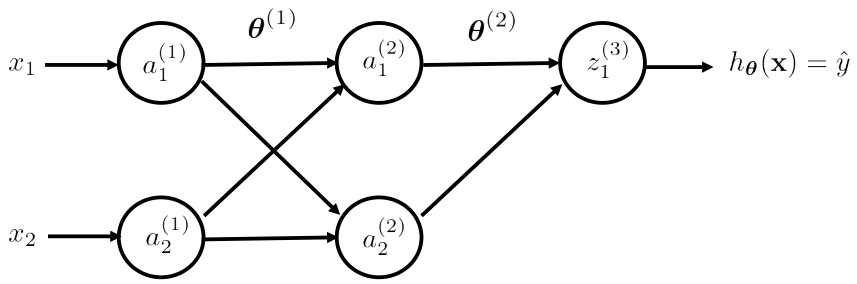


Figure 1: Neural network with 3 layers.

Define

$$\mathbf{a}^{(1)} = \mathbf{x} = (x_1, x_2), \mathbf{a}^{(2)} = g(\mathbf{z}^{(1)}), \text{ and } \mathbf{z}^{(l+1)} = \boldsymbol{\theta}^{(l)}\mathbf{a}^{(l)},\ l = 1, 2,$$

and the predicted value $\hat{y}$ is given by $h_{\boldsymbol{\theta}}(\mathbf{x}) = z_1^{(3)}$. Suppose that the squared error loss is used.

(a) Write a code that uses backpropagation for training the given neural network.

(b) Consider a single sample with $\mathbf{x} = [2\ 1]^T$ and $y = 3$. Initialize all weights of the edges using independent and identically distributed (i.i.d.) uniform(0,1) random variables (i.e., the probability that it belongs to the interval (0, $p$) is equal to $p$, $p \in (0, 1)$). Train the neural network using the sample for 50 iterations with a learning rate $\gamma = 0.05$. Plot the squared error loss as a function of iteration.