1. Consider the following multilayer perceptron.
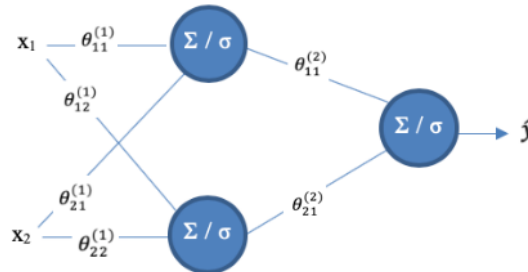
The network should implement the XOR function. Perform one epoch of *backpropagation* as introduced in the lecture on multilayer perceptrons.

Notes:

- The activation function f for a perceptron is the *sigmoid function*:

$$f(x) = \frac{1}{1 + e^{-x}}$$

- The thresholds are not shown in the network. The threshold nodes are set to -1.

- Use the following initial parameter values:

$$\theta_{01}^{(1)} = 2 \qquad \theta_{02}^{(1)} = -1 \qquad \theta_{01}^{(2)} = -2$$
$$\theta_{11}^{(1)} = 6 \qquad \theta_{12}^{(1)} = 8 \qquad \theta_{11}^{(2)} = 6$$
$$\theta_{21}^{(1)} = -6 \qquad \theta_{22}^{(1)} = -8 \qquad \theta_{21}^{(2)} = -6$$

- The learning rate is set to $\eta = 0.7$

i. Compute the activations of the hidden and output neurons.

ii. Compute the error of the network.

iii. Backpropagate the error to determine $\Delta\theta_{ij}$ for all weights θ_{ij} and updates the weight θ_{ij}.

2. What is the difference between "model bias" and "model variance"?

i. Why is a high bias, low variance classifier undesirable?

ii. Why is a low bias, high variance classifier (usually) undesirable?

3. Describe how validation set, and cross-validation can help reduce overfitting?