Large Scale Machine Learning and Deep Learning
Review Questions 4

1. Why adding more neurons to a single layer neural network cannot solve the XOR problem, but adding more layers can?

   _____

2. Assume a feedforward neural network with one hidden layer, in which the output of the hidden units and output units are computed by functions $\mathbf{h} = \mathbf{f}(\mathbf{x})$ and $\mathbf{out} = \mathbf{g}(\mathbf{h})$, respectively. Show that if we use linear functions in $\mathbf{f}$ and $\mathbf{g}$, e.g., $\mathbf{h} = \mathbf{f}(\mathbf{x}) = \mathbf{w}_1^\mathsf{T}\mathbf{x}$ and $\mathbf{out} = \mathbf{g}(\mathbf{h}) = \mathbf{w}_2^\mathsf{T}\mathbf{h}$, then the feedforward network as a whole would remain a linear function of its input.

   _____

3. What's the problem of using `step` function as an activation function in deep feedforward neural networks?

   _____

4. Compute the value of $\mathbf{w}_2$ and $\mathbf{w}_8$ after the first iteration of the backpropagation in the following figure. Assume all the neurons use the ReLU activation function and we use squared error function as the cost function. In this figure, red and orange colors indicate the initial values of the weights and biases, while the numbers in blue show the input and true output values.