

---

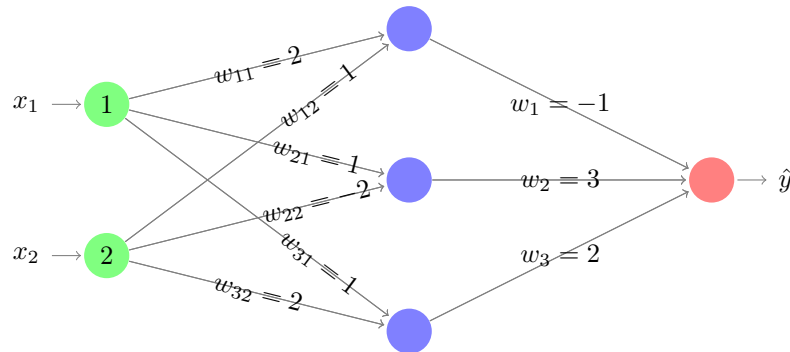
# Tutorial Questions | Week 4

## COSC2779 - Deep Learning

---

This tutorial is aimed at reviewing optimization techniques and regularization in deep learning. Please try the questions before you join the session.

1. Calculate the gradient update for the neural network below with given initialized weights, if the input data point is  $\mathbf{x} = [1, 2]$  and the expected output is 3. The loss is mean squared error, hidden units have ReLU activation, linier activation in output unit and the learning rate is 0.1. Initial biases are all zero.



2. What problem(s) will result from using a learning rate that's too  
(a) high?  
(b) low?
3. What is a saddle point? What is the advantage/disadvantage of Stochastic Gradient Descent in dealing with saddle points?
4. One of the difficulties with the sigmoid activation function is that of saturated units. Briefly explain the problem, and whether switching to tanh fixes the problem
5. What happens if we use batch-norm after an affine layer with bias? what effect does increasing the bias has?
6. Explain why dropout in a neural network acts as a regularizer.
7. Why do we often refer to L2-regularization as “weight decay”? Derive a mathematical expression to explain your point.
8. Explain what effect will the following operations generally have on the bias and variance of your model.  
(a) Regularizing the weights  
(b) Increasing the width of the layers  
(c) Using dropout to train a deep neural network  
(d) Getting more training data