## **Why do we care about the chain rule of derivatives**

Importance of chain rule in Deep Learning

1. Let us look at a sample chain rule flow of a shallow neural network
2. Here, the output ŷ is a composite dependent on input x and all of the parameters w
3. Now, for the gradient, we want the derivative of the loss function with respect to the various weights
4. If we want the derivative w.r.t w2 then we do the following
5. Here, computation happens from input layer to the output layer ie forward propagation
6. Derivative calculation happens backwards from the output layer to the input, ie back propagation