**Simple Illustration of Programmable Backpropagation**

Lei Mao

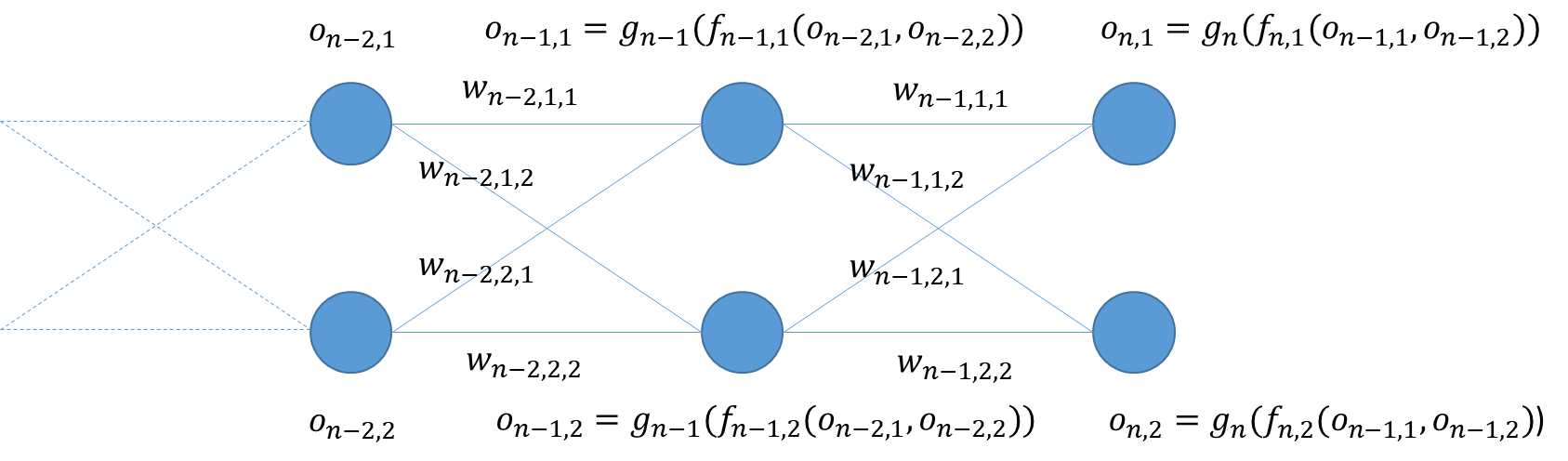
Department of Computer Science

University of Chicago

5/17/2017

**Introduction**

Suppose this is our neural network. I am only showing the last three layers of it.



where is the weights connecting different neurons and is the output from the neuron.

where function is the activation function at layer .

Suppose the target value for on the last layer is , we could compute the error using a certain loss function:

Then we could backpropagate this error to each weights at each layer. The main idea of back propagation is multivariable chain rule. We will calculate the derivatives at the last layer , and used the derivatives calculated in the layer to calculate derivatives in the layer , and so on. The backpropagated errors from one layer to another are colored using different colors. We will see that this backpropagation follows certain rules and is totally programmable.

**Backpropagation**

**Calculate error derivatives**

**Layer**

**Calculate (you don’t even have to do some of them)**

**Further calculate**

**Weight derivatives in layer**

**Derivatives used in layer**

**Layer**

**Calculate (you don’t even have to do some of them)**

**Further calculate**

**Weight derivatives in layer**

**Derivatives used in layer (multivariable chain rule)**

**Layer**

**Use the same way to calculate derivatives…**

**Weight derivatives in layer**

**Derivatives used in layer (multivariable chain rule)**

**Summary**

We can see from the above derivation, it could be a simple iterative computer program.

Basically, what we are doing in the backpropagation program is that, at each layer , we calculate the derivatives of activation function at each neuron, with a combination of the weights and outputs from the previous layer , and the propagated errors of layer which was already calculated at layer .

I am personally not interested in writing a program of doing backpropagations or autograds, since there are already tons of them. But this material unveil the programmable nature of backpropagation while most of the other materials I read previous did not have any concept of programming when they were introducing backpropagation. They usually just provide some simple chain rule examples without even providing a real neuron network. So the reader could only have some idea that the backpropagation is about the chain rules. But for programmers, we still do not quite understand such complicated derivations could be done in simple computer programs. The backpropagation program then become a black box even we are using it every day.