Introduction to Deep Learning - Week 2

<http://playground.tensorflow.org> for an online platform to compose MLPs

To train an MLP we need to calculate the hidden layer hyper-parameters automatically and fast.

The graph of derivatives is the MLP graph with reversed arrows. The chain rule is the sum of all the paths on this diagram. If you use the chain rule to expand all the derivatives then each term will relate to one backwards path through the MLP.

To use SGD we use a loss function on the sigmoid-ed output of the neuron.

Backpropagation uses the fact that the later derivatives in an MLP are reused as you go backwards in the network, meaning the value of each node derivative only has to be calculated once. First you make a forward-pass to determine how the node transforms its inputs and get the values at that point.

Matrix multiplication can be done fast in comparison to loops, and done on both CPU (e.g. BLAS) and GPU (e.g. cuBLAS).

Sjonni Notes