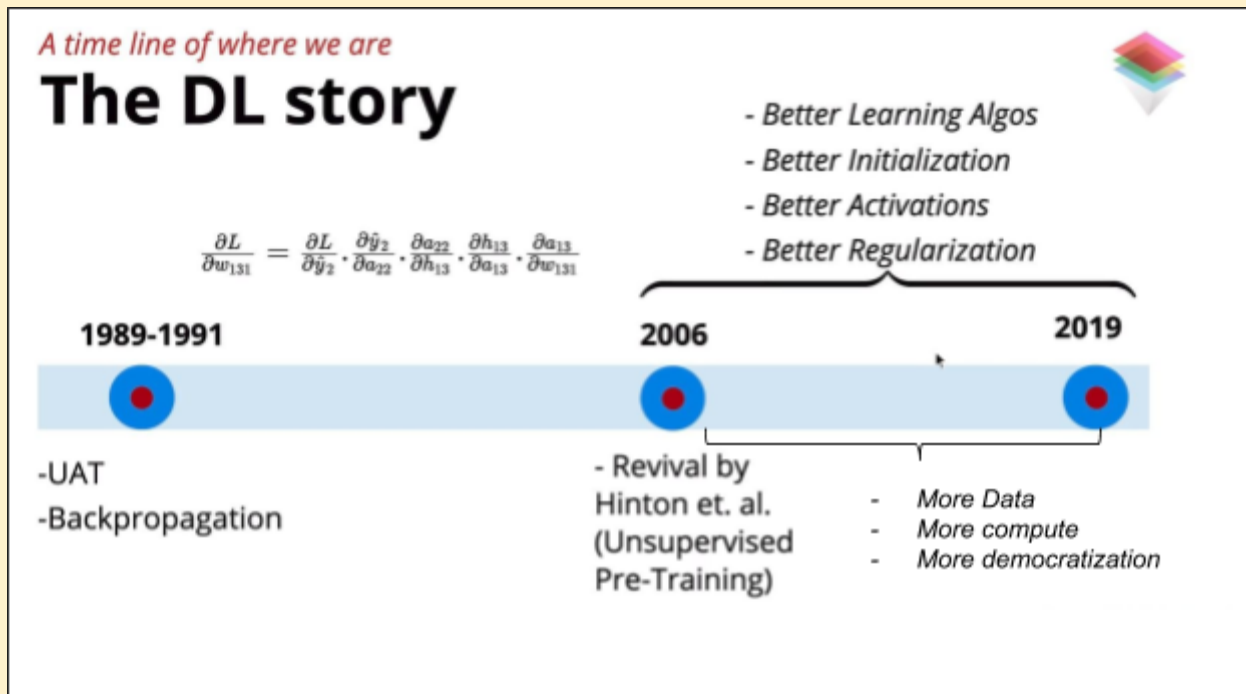


Optimization Algorithms

A quick history of DL to set the context

1. The following illustration shows the progress of Deep Learning over the last 3 decades



2. Some of the salient points in the DL-timeline are as follows
 - a. **1989-1991**
 - i. Universal Approximation Theorem: we will be able to approximate any kind of function with our Neural Network
 - ii. Backpropagation: Derivative calculation happens backwards from the output layer to the input, ie back propagation. It is nothing but Gradient Descent(1847) applied with the chain rule
 - b. **1993-1994**
 - i. A lot of work was done on Recurrent Neural Networks
 - c. **1998**
 - i. LSTMs (Long Short-Term Memory) were proposed
 - ii. Work done on Convolutional Neural Networks
 - d. **2006**
 - i. Revival of DL by Hinton et. al. with the proposal of Unsupervised Pre-training
 - ii. People's interest in DL started increasing.
 - e. **2019**
 - i. Better Learning Algorithms, Initializations, Activation and Regularization
 - ii. More Data, compute and democratization.