## Vanishing gradient Problem

## Decay of information through time

Vanishing Gradient Problem occurs when we try to train a Neural Network model using Gradient based optimization techniques. What happens is that as we keep on adding more and more Hidden layers in the model, the learning speed of the next hidden layers in the model keep on getting faster and faster.

Generally, adding more hidden layers tends to make the network able to learn more complex arbitrary functions, and thus do a better job in predicting future outcomes. This is where Deep Learning is making a big difference due to the thousands and millions of hidden layers it has , we can now make sense of highly complicated data such as images , speeches , videos etc. and do Speech Recognition and Image Classification , Image Captioning etc.

Now when we do Back-propagation i.e. moving backward in the Network and calculating gradients of loss (Error) with respect to the weights, the gradients tends to get smaller and smaller as we keep on moving backward in the Network. This means that the neurons in the Earlier layers learn very slowly as compared to the neurons in the later layers in the Hierarchy. The Earlier layers in the network are slowest to train.

**Earlier layers** in the Network are important because they are responsible to learn and detecting the simple patterns and are actually the building blocks of our Network. Obviously, if they give improper and inaccurate results, then how can we expect the next layers and the complete Network to perform nicely and produce accurate results?

Due to vanishing gradient problem the Training process takes too long and the Prediction Accuracy of the Model will decrease.