The term deep learning refers to training of neural networks, sometimes very deep neural networks.

What exactly is a neural network?

Taking the example of house prediction, uaing linear regression we fit a straight line between the relative data points of price vs size of the house, so neural network can also be thought of as the same as linear regression.

Simple neuron inputs the size, computes the linear function and then finds out the output price.

The function which keeps negative as zero and for positives it is y=x, this function is relu function and stands for Rectified Linear Unit.

A larger neural network is then formed by taking many of these neurons and then stacking them together.

So let’s say our house prediction depends upon a variety of things, like size, number of bedrooms, zip code, wealth , how much a person is willing to pay for the house , well how much really matters to them. Size and bedroom responsible for family size, zip code responsible for walkability and zip code and wealth responsible for school. Al the 3 factors would then find out the price.

To manage the neural network, we just give the input X and output Y for a number of inputs in the training set. Given the input features the job of the neural network would be to predict the price Y. Neurons in the middle are called hidden neurons which take up input from all of the four input features.

Given enough data about X and Y neural networks are remarkably good at functions that map X to Y.

In supervised learning you have some input X and to find out out somr function mapping of some output Y.

Applications include

* Ad user info
* Image recognition computer vision
* Audio text transcript, send in an audio and get the output as text transcript.
* Machine / language translation
* Autonomous driving

Different types of neural networks are useful for different types of neural networks. Housing predictions are standard types of neural networks, for images there are convolutional neural network(CNN), audio has a temporal component, played out mostly over time, represented as 1-D time series, so for sequence data we sue RNN, for autonomous driving we use complex hybrid neural network architecture.

Application of supervised learning to structured and unstructured data, structural data basically means databases of data, where the parameters have a very well defined meaning.

Unstructured data include raw audio, images or text. Thanks to neural networks computers are now much better at understanding unstructured data.

In traditional machine learning techniques like SVM or logistic regression we see the accuracy increases for a short time with increasing data and then becomes constant.

Thanks to digitization of the society the amount of data increased exponentially, so increasing the data and increasing the depth of the neural network the performance keeps getting better and better.

Most reliable way to increase the performance is train a bigger network or throw more data on the neural network. Problems would be we run out of data or create a neural network too big to train.

In the small training sets the relative ordering of the network vs traditional ML approaches is not well defined. Performance depends much more on our skills at engineering features.

In the big data regime , we see very large neural nets perform very well than traditional ML approaches.

Algorithmic innovations have been about trying to run the neural networks run much faster.

Big breakthrough is transfer from sigmoid function to Relu function. In case of sigmoid functions there are places where the gradient becomes essentially zero, so learning becomes very slow, as parameters change very slowly.

Advantage of Relu is gradient is one so it actually less likely to shrink to zero. So gradient descent works much faster.