Neural Networks and Deep Learning summary –

In this section we cover the guide to Deep Learning using matlab.

By reading Chapter 1 & 6 of “Neural Networks and Deep Learning” by Michael Nielsen

Also reading a paper ,

(LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-4441)

Machine learning systems power the way for Deep learning and Neural Networks to be formed. Representing a set of methods that allow a machine to be fed with raw data and to automatically discover the patterns , representations. Deep learning builds on that by representing learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level(starting with the raw input into a representation at a higher, slightly more abstract level.

* For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations.
* The key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure.
* It has been shown to be very good at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business, and government.
* Deep learning has produced extremely promising results for various tasks in natural language understating, particularly topic classification, sentiment analysis, question answering and language translation.

**RNN,** once unfolded in time, can be seen as very deep feedforward networks In which all the layers share the same weights. Using LSTM (long short-term memory) you can augment the network with explicit memory, so that you can learn long-term dependencies.

In a recurrent neural network , the backpropagation algorithm can be directly applied to the computational graph of the unfolded network, to compute the derivative of a total error.

To properly adjust the weight vector , the learning algorithm computes a gradient vector that for each weight , indicates by what amount the error would increase or decrease if the weight were increased by a tiny amount

The unsupervised deep learning area is being revived due to the fact that human and animal learning is largely unsupervised; we discover the structure of the world by observing it, not by being told the name of every object.

*Ultimately, major progress in artificial intelligence will come about through systems that combine representation learning with complex reasoning.*

A **multilayer neural network** , the chain rule of derivatives tells us how small effects are composed, the equations used tells us for computing the forward pass in a neural net with hidden layers and an output layer, each constitute a module through which one can back propagate gradients, and then there are other equations used for computing the backward pass; at each hidden layer we can compute the error derivative with respect to the output of each unit, which is the sum of the error derivatives with respect to the total inputs to the units in the layer above. Last but not least at the output layer, the error derivative with respect to the output of a unit is computed by differentiating the cost function.

The key advantage of deep learning is by automatically learning good features using a general-purpose learning procedure. A deep learning architecture is a multiplayer stack of simple modules, all (or most) of which are subject to learning, and many of which computer non-linear input- output mappings.

With a depth of 5-20, a system can implement extremely intricate functions of its inputs that are simultaneously sensitive to minute details and insensitive to irrelevant variations.

The backpropagation procedure to compute the gradient of an objective function with respect to the weights of a multilayer stack of modules is nothing more than a practical application of the chain rule for derivatives.

Many applications of deep learning use feedforward neural network architectures, which learn to map a fixed-size input to a fixed-size output.

At present the most popular non-linear function is the rectified linear unit (ReLU), which is the half-wave rectifier f(z) = max(z, 0).

The ReLU learns much faster in networks with many layers, allowing training of a deep supervised network without unsupervised pre-training.

Interest in deep feedforward networks was revived around 2006, The first major application of this pre-training approach was in speech recognition, and it was made possible by the advent of fast GPUs, allowing researchers to train networks 10 or 20 times faster.

The convolutional neural network (ConvNet) is a particular deep, feedforward network that was the most easiest to train and generalize much better than networks with full connectivity between adjacent layers.

ConvNet, are designed to process data that come in the form of multiple arrays, there are four key ideas behind ConvNets that take advantage of the properties of natural signals:

* local connections
* shared weights
* pooling and the use of many layers.

Its architecture structured as a series of stages. The first few stages are composed of two types of layers: convolutional layers and pooling layers.

Backpropagation gradients through a ConvNet is a simple as through a regular deep network, allowing all the weights in all the filter banks to be trained.

Deep neural networks exploit the property that many natural signals are compositional hierarchies, in which higher-level features are obtained by composing lower-level ones.

ConvNets have been applied with great success to the detection, segmentation and recognition of objects and regions in images.

When deep convolutional networks were applied to a data set of about a million images from the web that contained 1,000 different classes, they achieved spectacular results, almost halving the error rates of the best competing approaches. This success came from the efficient use of the GPUs, ReLUs, a new regularization technique called dropout and techniques to generate more training examples by deforming the existing ones. Recent ConvNet architectures have 10-20 layers of ReLUs, hundreds of millions of weights, and billions of connections between units. Training such large networks could have taken weeks but now in the progress in hardware, software, and algorithm parallelization have reduced training times to a few hours.

Deep-Learning theory shows that deep nets have two different exponential advantages over classic learning algorithms that do not use distributed representations. Having the power of composition and depending on the underlying data-generating distribution having an appropriate componential structure.

1st learning distributed representations enable generalization to new combinations of the values of learned features beyond those seen during training

2nd composing layers of representation in a deep net brings the potential for the hidden layers of a multilayer neural network which learn to represent the network’s inputs in a way that makes it easy to predict the target outputs.

RNNs are very powerful dynamic systems, but training them has proved to be problematic because the back propagated gradients either grow or shrink at each time step, so many times steps they typically explode or vanish.