DEEP LEARNING REVIEW

Yann LeCun, Yoshua Bengio & Geoffrey Hinton
Nature 2015

-Presented by Divya Chitimalla

What is deep learning

- Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.
- Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer.
- Key aspect of deep learning is that layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure
- Breakthroughs in learning networks, processing images, video, speech and audio.

Conventional ML vs. Deep learning

- Conventional machine-learning (ML) techniques are limited in their ability to process natural data in their raw form.
- It **requires careful engineering** and considerable domain expertise to design a feature extractor that transformed the raw data into a suitable internal representation or feature
- Representation learning is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification.
- Deep-learning methods **are representation-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.
- With the composition of enough such transformations, very complex functions can be learned.
- For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and **suppress irrelevant variations**.

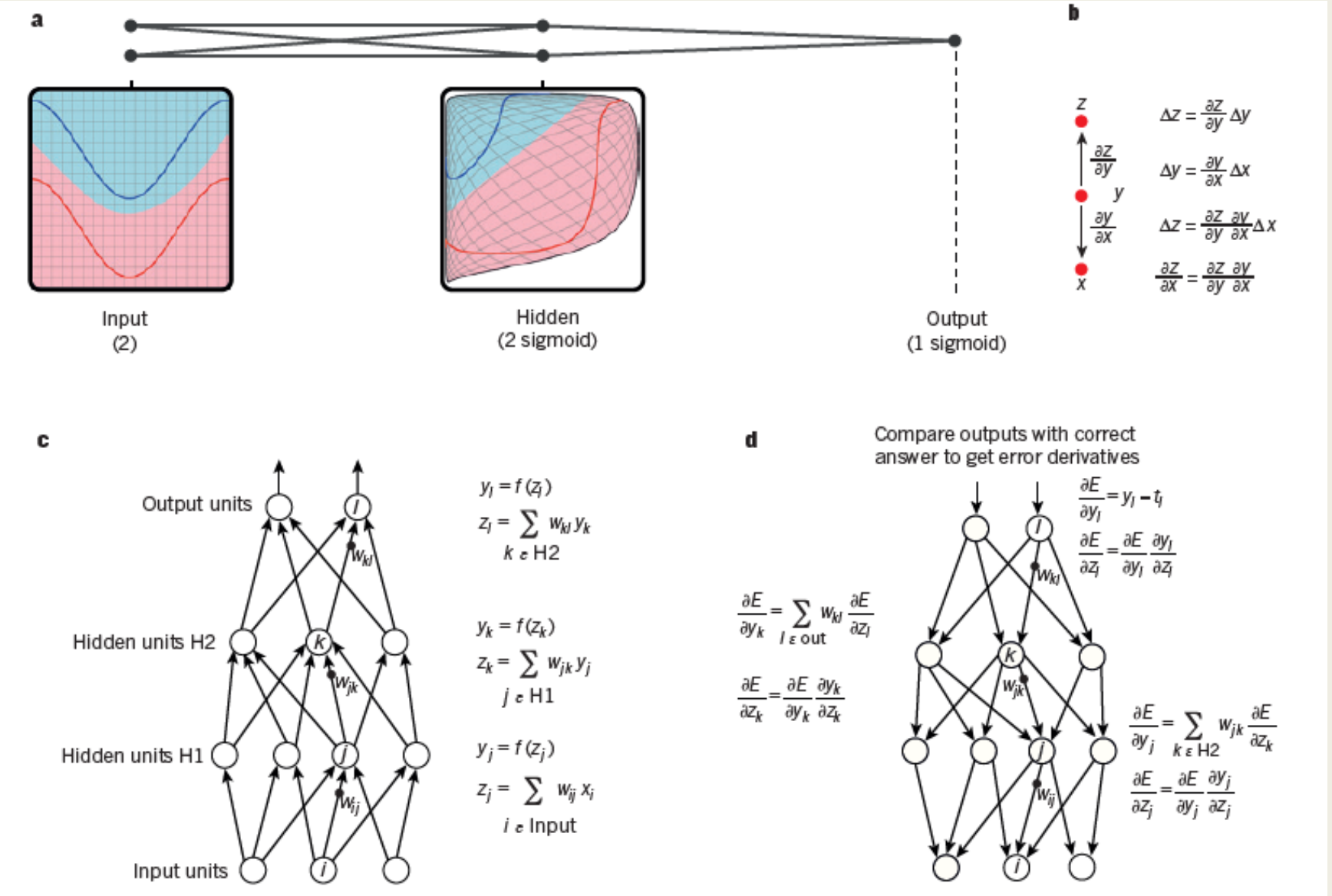
Supervised learning

- Most common form of machine learning
- System that can be classified using labels using training, compute an objective function that measures the error (or distance) between the output scores and the desired pattern of scores and then modifies its internal adjustable parameters to reduce this error.
- In deep-learning system, there are millions of these adjustable weights, and hundreds of millions of labelled examples with which to train the machine.
- 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 weight vector is then adjusted in the opposite direction to the gradient vector.
- **Stochastic gradient descent (SGD)** consists of showing the input vector for a few examples, computing the outputs and the errors, computing the average gradient for those examples, and adjusting the weights accordingly.

Neural network with back propagation

- Compute the gradient of an objective function with respect to the weights of a multilayer stack of modules using chain rule for derivatives.
- The key insight is that the derivative (or gradient) of the objective with respect to the input of a module can be computed by working backwards from the gradient with respect to the output of that module (or the input of the subsequent module)
- Backpropagation equation can be applied repeatedly to propagate gradients through all modules, starting from the output
- To go from one layer to the next, a set of units compute a weighted sum of their inputs from the previous layer and pass the result through a non-linear function. At present, the most popular non-linear function is the rectified linear unit (ReLU) $\tanh(z)$ or $1/(1 + \exp(-z)$

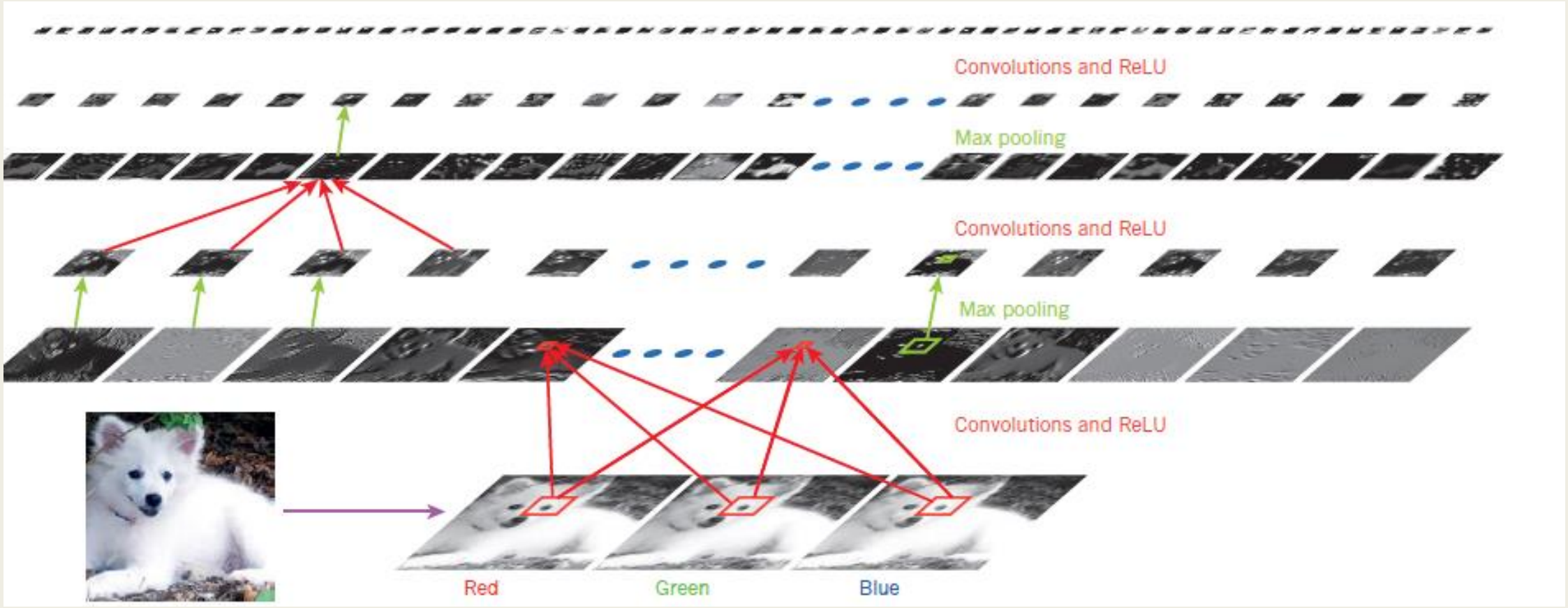
Neural network with back propagation



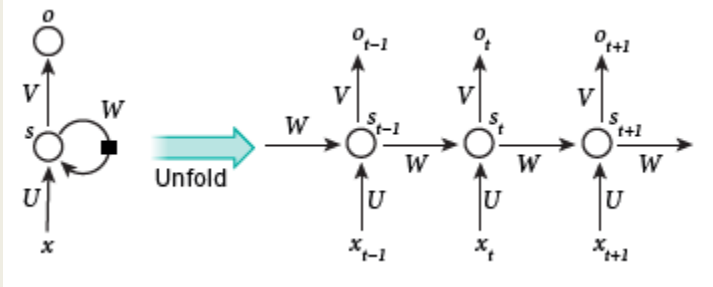
Convolution neural network

- ConvNets are designed to process data that come in the form of multiple arrays, for example a color image composed of three 2D arrays containing pixel intensities.
- The architecture of a typical ConvNet is structured as a series of stages. The first few stages are composed of two types of layers: **convolutional layers and pooling layers**.
- Units in a **convolutional layer are organized in feature maps**, within which each unit is connected to local patches in the feature maps of the previous layer through a set of weights called a filter bank.
- The result of this local weighted sum is then passed through a **non-linearity such as a ReLU**.
- All units in a feature map share the same filter bank. Different feature maps in a layer use different filter banks.
- First, in array data such as images, **local groups of values are often highly correlated, forming distinctive local motifs** that are easily detected. Second, the local statistics of images and other **signals are invariant to location**.
- Typically used in image recognition, speech processing etc.

Convolution neural network



Recurrent neural networks



- RNNs process an input sequence one element at a time, maintaining in their hidden units a ‘state vector’ that implicitly contains information about the history of all the past elements of the sequence. When we consider the outputs of the hidden units at different discrete time steps as if they were the outputs of different neurons in a deep multilayer network
- RNNs are very powerful dynamic systems, but training them has proved to be problematic because the backpropagated gradients either grow or shrink at each time step, so over many time steps they typically explode or vanis
- RNNs can be seen as very deep feedforward networks in which all the layers share the same weights. Although their main purpose is to learn long-term dependencies, theoretical and empirical evidence shows that it is difficult to learn to store information for very long.
- To correct for that, one idea is to augment the network with an explicit memory. The first proposal of this kind is the long short-term memory (LSTM) networks that use special hidden units, the natural behaviour of which is to remember inputs for a long time

Future of deep learning

- It is everywhere
- Unsupervised learning will become far more important in the longer term.
- 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.