Deep learning

Deep learning is a set of representation-learning methods, the input/raw data is automatically transformed with a general-purpose learning procedure to higher layers of abstraction. It helps to get rid of motifs and keep track of general features of something so that knowledge can be later used to classify or detect patterns.

Supervised learning

This technique requires a large data set of labeled information.

The way supervised learning works is that given the large dataset, it is trained by classifying the training data and then comparing with the real output so it can perform a gradient descent in the weights that define the *input* → *output* function taking the algorithm closer to a minimum where the output error is low on average.

After training, the algorithm is presented test data to evaluate the generalization ability of the machine.

Backpropagation to train multilayer architectures

Backpropagation consists of a series of passes that go adapting the layers' weights to tune the outputs.

This method was used when unsupervised learning showed up in 2006.

Convolutional Neural Networks - CNN

Mathematically, by using filter banks between layers, a discrete convolution is performed, hence the name.

The four main stages of these kind of NN are:

- Local connections. Many problems involve compositional hierarchies (letters → words → sentences), and this is considered by the NN.

- Shared weights. When working with data, some motifs can show up several times, that’s why the same filter is used in different stages.
- Pooling. This allows to identify patterns disregarding changes in position and appearance (let’s say in images).
- Many layers

Image understanding with deep convolutional networks

ConvNets where mainly forsaken until the ImageNet competition in 2012 where the outcomes had astonishing results.

This boom + the progress in hardware and software has caused that most big tech companies initate research and development projects.

Distributed representations and language processing

These kind of networks has shown to be really good with language processing as well, and they’re proven to be better than other traditional algorithms.

There was a debate in 2015 on which representation should be used: a logic-based or a neural-network-based one.

Recurring neural networks - RNN

When backpropagation was first introduced, it was used to train RNNs. These kind of networks process sequential data one step at a time but holding a state vector that keeps in context the processing.

Even though these nn have shown to be problematic when training, they have been used in NLP. The cases are prediction of next characters of a word, next words of a sentence and even more impressing: given a sentence in English, it ‘keeps the idea’ to output a probability distribution of matching words in French, which has quickly battled the state-of-the-art and raises serious doubts.

A RNN was meant to learn long-term dependencies but it was complicated, so the workaround was to add explicit memory so the information can be stored for longer. Other approaches regarding memory has had excellent results on question-answer scenarios.

The future of deep learning

The authors rely the future of deep learning on unsupervised one, since that’s the way humans and animals learn.

Many of these techniques are still in their infancy but they’re already producing impressive outcomes, but some work still needs to be done. Techniques on NLP, for example.

Extra concepts

SGD - Stochastic gradient descent

It’s a variation of the classical gradient descent but it’s performed several times over small samples of the original gradients. It is surprisingly good at it.

Surprisingly, local minima are not commonly an issue with big networks.

Facts on units

Near 2012, a new concept showed up called **dropout.** This proposes turning off random nodes.

In 2015, the most popular non-linear function was the rectified linear unit: ReLU.

Units that aren’t the input or output are usually called *hidden units.*