Notes - Deep Learning Three Giant's Survey

Friday, January 4, 2019 3:41 PM

How deep learning discovers the intricate structure in large datasets?
DL discovers the intricate structure in large datasets by using the back propagation 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 previous layer.

What are the drawbacks of Conventional Machine Learning Techniques?
Earlier to representational learning and Deep networks, constructing a pattern-recognition or machine-learning system required careful engineering and considerable domain expertise to design a feature extractor that transformed the raw data (pixel values of an image) into a suitable internal representation or feature vector from which the classifier could detect or classify patterns in the input.

What is representational learning?
It 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.

What are Deep Learning Methods and how they discover the representations?
Deep learning methods are an example of representational learning. They have multiple levels of representation obtained by composing simple but non-linear modules that each transform the representation at one level into a representation at a higher, slightly more abstract level. With the composition of enough such transformation complex functions can be learned.

How the higher layers in deep learning methods work in the classification tasks?
Higher layers of representation in the deep learning model amplify aspects that are important for discrimination and at the same time suppress irrelevant variations.

What is a good example of representational learning?
An image is in the form of an array of pixel values and the learned features in the first layer of representation typically represents the presence or absence of edges at particular orientations and locations in the image.
The second layer typically detects the pattern by spotting particular arrangements of edges, regardless of small variations in the edge positions.
The third layer arranges these patterns into larger combinations that belongs to the parts of familiar objects and subsequent layers would detect objects as combinations of these parts.
The key aspect is that these layers are learned from data using a general purpose learning procedure and they are not designed by human engineers.

What is Supervised Learning?
In this class of machine learning, the algorithms are presented with the dataset of images and relevant labels for each image. During training the machine is shown an input image and produce an output in the form of vector of scores, one for each category. The expectation is to have highest score to the desired category. But it is high likely to happen during the training. Then an objective function will measure the error between the output scores and desired pattern of scores. The machine then modifies the adjustable internal parameters and reduce this error. These parameters are called as weights which are real numbers behave as the knob that defines the input-output function of the machine.

How the parameters of the algorithm are changed to minimize the error from the objective function?
The parameters of the algorithm have to be changed properly in order to minimize the error efficiently. The learning algorithm computes a gradient that, for each weight, indicates by what amount the error would increase or decrease if the weight were increased by the tiny amount. Then the weight vector is

adjusted in the direction opposite to the gradient vector.

What is meant by Stochastic Gradient Descent (SGD) method?
In SGD, the machine is shown the input vector for few examples, computing the output and errors, computing the average gradient for those examples and adjust the weights accordingly. The process is repeated for many small sets of examples from the training set until the average of the objective function stops decreasing. The term is called stochastic because each small set of examples gives a noisy estimate of the average gradient over all examples. This simple procedure usually finds a good set of weights quickly when compared with other optimization techniques which are elaborate and complex.

What is the advantage of deep learning over the classical linear classifier?
Problems such as image and speech-recognition require the input-output function to be insensitive to irrelevant variations of the input, such as variations in position, orientation or illumination of an object, or variations in the pitch or accent of speech, while being very sensitive to particular miniature variations (the difference between a white wolf and a breed of wolf-like white dog called a Samoyed).
For example:
At pixel level, images of two Samoyeds in different poses and in different environments may be very different from each other. While images of a Samoyed and a Wolf in the same position and on similar backgrounds may be very similar to each other.
A linear classifier or any other shallow classifier operating on raw pixels would not be able to distinguish the latter two and put the former two into same category. A good feature extractor is required to solve this selectivity-invariance dilemma. This feature extractor produces representations that are selective to the aspects of the image that are important for discrimination, but that are invariant to irrelevant aspects such as pose of the animal.
But deep learning offers a general-purpose learning procedure that can help to learn good features automatically.

How deep learning solves the selectivity-invariance dilemma?
A DL architecture is a multilayer stack of simple modules. They are subject to learning and compute non-linear input-output mappings. Each module in the stack transforms its input to increase both the selectivity and invariance of the representation. With such multiple non-linear layers the model can implement extremely intricate functions of its inputs that are simultaneously sensitive to minute details and insensitive to large irrelevant variations such as background, pose, lighting and surrounding objects