Chapter 8
Deep Belief Networks (DBNs)

Abstract. This chapter covers successful applications in deep learning with remarkable capability to generate sophisticated and invariant features from raw input signal data. New insights of the visual cortex and studies in the relations between the connectivity found in the brain and mechanisms for mind inference have enlightened the development of deep neural networks. In this chapter the motivation for the design of these architectures points out towards models with many layers exhibiting complex behavior enhanced by thousands of neurons in each layer. By contrast, shallow neural networks have admittedly less capability with respect to inference mechanisms since no feature detectors are found in their hidden layer. Then the chapter formalizes Restricted Boltzmann Machines (RBMs) and Deep Belief Networks (DBNs), which are generative models that along with an unsupervised greedy learning algorithm CD–k are able to attain deep learning of objects. A robust learning adaptive size method is presented. Moreover, a GPU parallel implementation yields high speedups as compared to conventional implementations. Results in benchmark data sets are presented as well as further discussion of these models.

8.1 Introduction

Recent empirical and theoretical advances in deep learning methods have led to a widespread enthusiasm in the pattern recognition and ML areas [148,110]. Inspired by the depth structure of the brain, deep learning architectures encompass the promise of revolutionizing and widening the range of tasks performed by computers [148]. In recent months deep learning applications have been growing both in number and accuracy [148]. State-of-the-art technologies such as the Apple’s Siri personal assistant or Google’s Street View already integrate deep NNs into their systems, asserting their potential to increase the specter of automated systems capable of performing tasks that would otherwise require humans [148]. Moreover, just a few months ago, a team of graduate students of Geoffrey E. Hinton won the top prize in a contest aimed at finding molecules that might lead to new drugs. This
was a particularly impressive achievement because never before had a deep learning architecture based-system won a similar competition and the software was designed with no prior knowledge on how the molecules bind to their targets, using only a relatively small dataset \[148\].

Deep architecture models reflect the results of many levels of composition of non-linear operations in their outputs \[110, 192, 14\]. The idea is to have feature detector units at each layer (level) that gradually extract and refine more sophisticated and invariant features from the original raw input signals. Lower layers aim at extracting simple features that are then clamped into higher layers, which in turn detect more complex features \[113\]. In contrast, shallow models (e.g. linear models, one hidden layer NNs, SVMs) present very few layers of composition that basically map the original input features into a problem-specific feature space \[110, 251\]. Figure 8.1 illustrates the main architectural differences between deep and shallow models.

**Fig. 8.1** Deep architectures versus shallow ones