YOSHUA BENGIO, U. Montreal
On the Challenge of Learning Abstractions

We argue that learning to be intelligent involves the learning of highly varying functions, in a mathematical sense. We present results suggesting strongly that most currently popular statistical learning approaches to learning flexible functions have fundamental limitations that render them inappropriate for learning highly varying functions. The first issue concerns the representation of such functions with what we call shallow model architectures. We discuss limitations of shallow architectures, such as so-called kernel machines, boosting algorithms, decision trees, and one-hidden-layer artificial neural networks. Mathematical results in circuits complexity theory helps us understand the issue. The second issue is more focused and concerns kernel machines with a local (e.g. Gaussian) kernel.

We show that they have a limitation similar to those already proved for older non-parametric methods, and connected to the so-called curse of dimensionality. Though it has long been believed that efficient learning in deep architectures is difficult, recently proposed computational principles for learning in deep architectures may offer a breakthrough. An idea that emerges from the experiments performed with these algorithms is that in order to optimize a highly non-convex functions, humans and machines could be exploiting the pedagogical approach: learn simple concepts first, and when they are mastered use them to express and learn more abstract concepts. Selecting training examples and the order in which they are presented (just like teachers do with children) could be a way to guide this difficult optimization problem by solving a series of gradually more complex problems embedded in each other.