Shallow vs. Deep Sum-Product Networks

Yoshua Bengio and Olivier Delalleau

Dept. IRO, Université de Montréal
C.P. 6128, Montreal, Qc, H3C 3J7, Canada
yoshua.bengio@umontreal.ca, delallea@iro.umontreal.ca
http://www.iro.umontreal.ca/∼lisa

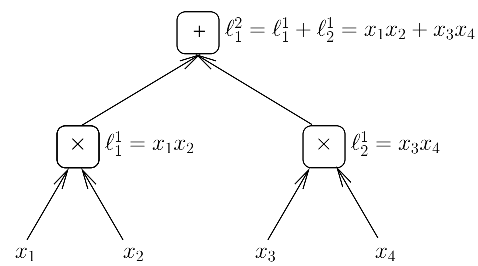

Figure 1: Basic binary sum-product network.

We investigate efficient representations of functions that can be written as outputs of so-called *sum-product networks*, that alternate layers of product and sum operations (see Fig 1 for a simple sum-product network). We find that there exist families of such functions that can be represented much more efficiently by deep sum-product networks (i.e. allowing multiple hidden layers), compared to shallow sum-product networks (constrained to using a single hidden layer). For instance, there is a family of functions f_n where n is the number of input variables, such that f_n can be computed with a deep sum-product network of $\log_2 n$ layers and $n-1$ units, while a shallow sum-product network (two layers) requires $2^{\sqrt{n}-1}$ units.

These mathematical results are in the same spirit as those by Håstad and Goldmann (1991) on the limitations of small depth computational circuits. They motivate using deep networks to be able to model complex functions more efficiently than with shallow networks. Exponential gains in terms of the number of parameters are quite significant in the context of statistical machine learning. Indeed, the number of training samples required to optimize a model's parameters without suffering from overfitting typically increases with the number of parameters. Deep networks thus offer a promising way to learn complex functions from limited data, even though parameter optimization may still be challenging.

References

Håstad, J. and Goldmann, M. (1991). On the power of small-depth threshold circuits. *Computational Complexity*, **1**, 113–129.