Neural Networks and (Virtual) Extended Formulations

Christoph Hertrich*[1] and Georg Loho[2,3]

[1]Université libre de Bruxelles – Belgique
[2]Freie Universität Berlin – Allemagne
[3]University of Twente – Pays-Bas

Résumé

Neural networks with piecewise linear activation functions, such as rectified linear units (ReLU) or maxout, are among the most fundamental models in modern machine learning. We make a step towards proving lower bounds on the size of such neural networks by linking their representative capabilities to the notion of the extension complexity xc(P) of a polytope P, a well-studied quantity in combinatorial optimization and polyhedral geometry. To this end, we propose the notion of virtual extension complexity vxc(P) = min { xc(Q)+xc(R) — P+Q=R }. This generalizes xc(P) and describes the number of inequalities needed to represent the linear optimization problem over P as a difference of two linear programs. We prove that vxc(P) is a lower bound on the size of a neural network that optimizes over P. While it remains open to derive strong lower bounds on virtual extension complexity, we show that powerful results on the ordinary extension complexity can be converted into lower bounds for monotone neural networks, that is, neural networks with only nonnegative weights. Furthermore, we show that one can efficiently optimize over a polytope P using a small virtual extended formulation. We therefore believe that virtual extension complexity deserves to be studied independently from neural networks, just like the ordinary extension complexity. As a first step in this direction, we derive an example showing that extension complexity can go down under Minkowski sum.

*Intervenant