CRUNCH Seminars at Brown, Division of Applied Mathematics

Monday – April 22, 2019

Deep vs. Shallow Networks: an Approximation Theory Perspective

Hrushikesh Mhaskar

The paper briefly reviews several recent results on hierarchical architectures for learning from examples, that may formally explain the conditions under which Deep Convolutional Neural Networks perform much better in function approximation problems than shallow, one-hidden layer architectures. The paper announces new results for a non-smooth activation function – the ReLU function – used in present-day neural networks, as well as for the Gaussian networks. We propose a new definition of relative dimension to encapsulate different notions of sparsity of a function class that can possibly be exploited by deep networks but not by shallow ones to drastically reduce the complexity required for approximation and learning.