Mathematics of Deep Learning

Rene Vidal, Johns Hopkins University

The past few years have seen a dramatic increase in the performance of recognition systems thanks to the introduction of deep networks for representation learning. However, the mathematical reasons for this success remain elusive. For example, a key issue is that the neural network training problem is non-convex, hence optimization algorithms may not return a global minima. In addition, the regularization properties of algorithms such as dropout remain poorly understood. The first part of this talk will overview recent work on the theory of deep learning that aims to understand how to design the network architecture, how to regularize the network weights, and how to guarantee global optimality. The second part of this talk will present sufficient conditions to guarantee that local minima are globally optimal and that a local descent strategy can reach a global minima from any initialization. Such conditions apply to problems in matrix factorization, tensor factorization and deep learning. The third part of this talk will present an analysis of the optimization and regularization properties of dropout for matrix factorization in the case of matrix factorization. Examples from neuroscience and computer vision will also be presented.

Rene Vidal is the Herschel L. Seder Professor of Biomedical Engineering and the Inaugural Director of the Mathematical Institute for Data Science at The Johns Hopkins University. His research focuses on the development of theory and algorithms for the analysis of complex high-dimensional datasets such as images, videos, time-series and biomedical data. Dr. Vidal has been Associate Editor of TPAMI and CVIU, Program Chair of ICCV and CVPR, co-author of the book "Generalized Principal Component Analysis"" (2016), and co-author of more than 200 articles in machine learning, computer vision, biomedical image analysis, hybrid systems, robotics and signal processing. He is a fellow of the IEEE, IAPR and Sloan Foundation, a ONR Young Investigator, and has received numerous awards for his work, including the 2012 J.K. Aggarwal Prize for ""outstanding contributions to generalized principal component analysis (GPCA) and subspace clustering in computer vision and pattern recognition"" as well as best paper awards in machine learning, computer vision, controls, and medical robotics.