



ICTS

INTERNATIONAL
CENTRE *for*
THEORETICAL
SCIENCES

TATA INSTITUTE OF FUNDAMENTAL RESEARCH

ICTS Special Seminar

Title : Stabilizing Gradients for Deep Neural Networks

Speaker : Inderjit S. Dhillon, University of Texas, Austin and Amazon

Date : Thursday, September 27, 2018

Time : 3:00 PM

Venue : Emmy Noether Seminar Room, ICTS Campus, Bangalore

Abstract : Deep Learning has shown great results in the difficult modern applications of computer vision, speech processing and natural language processing. In general, larger depth of the network has led to superior performance in these applications. However, there still remain many fundamental questions in deep learning. Among them, vanishing and exploding gradients are two main obstacles in training deep neural networks, especially in capturing long range dependencies. In this tutorial, I will explain the reason that gradients can vanish and explode as the depth becomes larger. This understanding leads to a natural solution that uses tools common in numerical linear algebra, specifically the use of elementary orthogonal transformations that allow us to work with a more stable representation of the weight matrices that arise in deep learning. This in turn allows us to control the norm of the gradients, thus leading to a solution of the exploding and vanishing gradient problem. I will illustrate that this gradient stabilization leads to better results on synthetic and real applications.