

The Information Theory of Deep Neural Networks and the Computational Benefits of the Hidden Layers

Prof. Naftali Tishby

*ELSC Center for Brain Sciences &
The Benin School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem, Israel*



In the past several years we have developed a comprehensive theory of large scale learning with Deep Neural Networks (DNN), when optimized with Stochastic Gradient Decent (SGD). The theory is built on three theoretical components: (1) Rethinking the standard (PAC like) distribution independent worst case generalization bounds - turning them to problem dependent typical (in the Information Theory sense) bounds that are independent of the model architecture. (2) The Information Plane theorem: For large scale typical learning the sample-complexity and accuracy tradeoff is characterized by only two numbers: the mutual information that the representation (a layer in the network) maintain on the input patterns, and the mutual information each layer has on the desired output label. The Information Theoretic optimal tradeoff between the encoder and decoder information values is given by the Information Bottleneck (IB) bound for the rule specific input-output distribution. (3) The layers of the DNN reach this optimal bound via standard SGD training, in high (input & layers) dimension.

In this talk, Professor Tishby will briefly review these results and discuss two new surprising outcomes of this theory: (1) The computational benefit of the hidden layers, (2) the emerging understanding of the features encoded by each layer which follows from the convergence to the IB bound.

Wednesday, May 29th, 2019 at 2:15 pm

**MPIDS, Prandtl lecture hall,
Am Faßberg 17, Göttingen**

**Max Planck Institute for Dynamics and Self-Organization
Living Matter Physics
Dr. Evelyn Tang**

Email: evelyn.tang@ds.mpg.de, Phone: +49-(0)551/5176-155
Am Faßberg 17, 37077 Göttingen, Germany