



## CSML and ECE SEMINAR

**Speaker:** Mehul Motani, National University of Singapore  
**Title:** Some Thoughts on Generalization in Deep Learning  
**Day:** June 14, 2022  
**Time:** 10:30am  
**Location:** B205 Engineering Quadrangle  
**Host:** H. Vincent Poor and Peter Ramadge

### Abstract:

A good learning algorithm is characterized primarily by its ability to predict beyond the training data, i.e., its ability to generalize. What makes a learning algorithm have the ability to generalize? And can we predict when a learning algorithm will generalize well? We believe that a clear answer to these questions is still elusive. In this talk, we will share some perspectives based on our work to understand generalization in deep learning algorithms. It is well accepted that the complexity of the classifier's function space is key to generalization. Recent work has shown that even within a classifier's function space, there can be significant differences in the ability to generalize. Motivated by this, we propose a new measure of complexity called Kolmogorov Growth (KG), which leads to new generalization error bounds that only depend on the final choice of the classification function. We also propose a novel method called network-to-network regularization which constrains the network trajectory to remain in the low KG zone during training. Minimizing KG while learning is akin to applying Occam's razor to neural networks and leads to clear improvements in generalization and robustness to label noise. To complement the complexity based approaches, we also explore an information theoretic perspective on generalization. Most research, such as the information bottleneck theory, studies the dynamics of mutual information (MI). However, estimating MI in high dimensions and in deterministic settings is problematic, often resulting in widely varying estimates for different estimators. In our work, we use a variant called sliced mutual information (SMI), which is scalable and efficient to compute. Unlike MI, the SMI between the features and labels can encode geometric properties of the feature distribution, making SMI relevant to the generalization error of the classifier.

### Bio:

Mehul Motani received the B.E. degree from Cooper Union, New York, NY, the M.S. degree from Syracuse University, Syracuse, NY, and the Ph.D. degree from Cornell University, Ithaca, NY, all in Electrical and Computer Engineering. Dr. Motani is currently an Associate Professor in the Electrical and Computer Engineering Department at the National University of Singapore (NUS) and a Visiting Research Collaborator at Princeton University. Previously, he was a Visiting Fellow at Princeton University. He was also a Research Scientist at the Institute for Infocomm Research in Singapore, for three years, and a Systems Engineer at Lockheed Martin in Syracuse, NY for over four years. His research interests include information and coding theory, machine learning, biomedical informatics, wireless and sensor networks, and the Internet-of-Things. Dr. Motani was the recipient of the Intel Foundation Fellowship for his Ph.D. research, the NUS Annual Teaching Excellence Award, the NUS Faculty of Engineering Innovative Teaching Award, and the NUS Faculty of Engineering Teaching Honours List Award. He actively participates in the Institute of Electrical and Electronics Engineers (IEEE) and the Association for Computing Machinery (ACM). He is a Fellow of the IEEE and has served as the Secretary of the IEEE Information Theory Society Board of Governors. He has served as an Associate Editor for both the IEEE Transactions on Information Theory and the IEEE Transactions on Communications. He has also served on the Organizing and Technical Program Committees of numerous IEEE and ACM conferences.

