Learnable Knowledge and Information in Deep Neural Networks

Abstract: Shannon famously declared that, for the purpose of communication, the semantic aspect of information is irrelevant. For the purpose of learning, however, the task defines what part of the data is informative. Also, information as defined in communication is a property of a distribution, but modern deep networks trained from data are deterministic maps trained on a given dataset, not a distribution of them. As a result, most bounds are vacuous and even defining, let alone computing and bounding information, presents non-trivial challenges. In this talk, I will describe a notion of information in a learned representation that is well-defined, can be computed for large-scale real-world models, and yields non-vacuous generalization bounds. I will then show how such a notion of information can be used to compute the complexity of a learning task and define a topology in the space of tasks, so we can compute how “far” two tasks are, and whether it is possible to “reach” one from another (transfer learning). Once we know how to compute it, measuring information during the training process sheds light on phenomena which have been observed in both biological and artificial, systems, such as irreversibility (critical learning periods), and forgetting, pointing to fundamental information processes that are independent of the medium, whether biological or artificial.

Joint work with Alessandro Achille

Bio: Stefano Soatto is Vice President for AI Labs at AWS and a professor of Computer Science at UCLA. He received his PhD in Control and Dynamical Systems from Caltech in 1996 and was a Postdoc at Harvard and an Assistant Professor of Biomedical Engineering and Electrical Engineering at Washington University. He is a recipient of the David Marr Prize, he is a Fellow of the IEEE and received a number of best paper awards for his work with his students and collaborators. He currently oversees research leading to AWS AI Services in the areas of Computer Vision, Document Processing, Natural Language Processing, Speech Processing, Forecasting, Personalization, DevOps and others.