

Title: PAC learning high dimensional distributions.

Speaker: Dr. Sutanu Gayen, Postdoctoral Research Fellow at the School of Computing, NUS Singapore

Date and Time: Tuesday, 10 August, 2021 (15:30 hrs)

Zoom: Online

Abstract:

In this talk, we will look at several fundamental questions about high dimensional distributions. I will start by introducing the PAC learning framework to formally study such questions. In particular, we look at three broad problems which lie at the foundations of machine learning in high dimensions. For each problem, we will give an algorithm that takes a near-optimal number of samples and has an efficient runtime.

In the first problem, we are getting samples from an unknown high-dimensional distribution, and our goal will be to approximately learn a tree structure that describes the distribution. We will establish tight sample complexity for this problem in two cases: agnostic, when the unknown distribution can be arbitrary, and realizable, when the distribution is itself tree-structured. En route, we will give a better and optimal estimator for testing whether two random variables are independent or not.

Next, we will be given a graph structure, and our goal will be to fit an unknown distribution on the given structure using samples. We will give an efficient algorithm with tight sample complexity for this problem.

Finally, we will look at an important problem from causal inference, which is learning interventional distributions from observational data. I will first give an algorithm for this problem for efficiently learning the joint distribution. In contrast, I will show that learning a marginal distribution is in general harder than the graph isomorphism problem, by connecting this problem to that of testing samplable distributions.

Speaker Bio:

Sutanu Gayen is a post-doctoral research fellow in the School of Computing, NUS Singapore, hosted by Prof. Arnab Bhattacharyya since August 2019. He completed his Ph.D. in the Department of Computer Science and Engineering, University of Nebraska-Lincoln, advised by Prof. Vinodchandran N. Variyam from August 2013 to June 2019. His research interest broadly lies in the foundations of machine learning.