Abstract
This talk addresses statistical inference with mixture and hierarchical models: efficiency of parameter estimation in finite mixtures, and scalable clustering of multilevel structured data. It is well-known that due to weak identifiability and singularity structures of latent variable models’ parameter space, the convergence behaviors of parameter estimation procedures for mixture models remain poorly understood. In the first part of the talk, we describe a general framework for characterizing impacts of weak identifiability and singularity structures on the convergence behaviors of the maximum likelihood estimator in finite mixture models. This allows us to resolve several open questions regarding popular models such as Gaussian and Gamma mixtures, as well as to explicate the behaviors of complex models such as mixtures of skew normal distributions.

In the second part of the talk, we address a clustering problem with multilevel structured data, with the goal of simultaneously clustering a collection of data groups and partitioning the data in each group. By exploiting optimal transport distance as a natural metric for distributions and a collection of distributions, we propose an optimization formulation that allows to discover the multilevel clustering structures in grouped data in an efficient way. We illustrate the performance of our clustering method in a number of application domains, including computer vision.

Speaker Biography
Nhat Ho is currently a fifth year PhD student in the Department of Statistics at the University of Michigan, Ann Arbor. He works under the supervision of Professor Long Nguyen and Professor Ya’cov Ritov in the Department of Statistics. Ho’s current research interests include theories and applications of mixture and hierarchical models, clustering analysis, Bayesian nonparametrics, robust statistics, and statistical machine learning.