

**University of California, Irvine
Joint Statistics/ML/AI Seminar**

Unsupervised Validation for Unsupervised Learning

**Marina Meila
Professor, Dept. of Statistics
University of Washington**

**Tuesday, Nov. 19, 2019
1 p.m., 2011 Bren Hall
(Bldg. #314 on campus map)**

Machine learning is many times faster than humans at finding patterns, yet the task of validating these as ``meaningful" is still left to the human expert or to further experiment. In this talk I will present three instances in which unsupervised learning tasks can be augmented with data driven validation.

In the case of clustering, I will demonstrate a new framework of "proving" that a clustering is approximately correct, that does not require a user to know anything about the data distribution. This framework has some similarities to PAC bounds in supervised learning; unlike PAC bounds, the bounds for clustering can be calculated exactly and can be of direct practical utility.

In the case of non-linear dimension reduction, I will address the following well known issues. The output of manifold learning algorithms distorts distances, angles and other geometric properties of the original data. Our contribution is a statistically founded method to estimate and then cancel out the distortions, thus effectively preserving the distances in the original data. This method is based on the notion of augmenting the algorithm output with a Riemannian metric, i.e. with the information that allows it to reconstruct the original geometry.

Furthermore, the abstract coordinates obtained by dimension reduction are often identified, by visual inspection, with interpretable properties of the data. The third and last part of the talk will describe a method to semi-automate this process. The human expert provides a dictionary of meaningful functions, and our algorithm selects a subset of these that can parametrize a manifold via an arbitrary smooth non-linear transformation.

Joint work with Dominique Perrault-Joncas, James McQueen, Yu-chia Chen, Samson Koelle, Hanyu Zhang.

For directions/parking information, please visit <https://uci.edu/visit/maps.php> and <http://www.ics.uci.edu/about/visit/index.php>