



Working Group on Statistical Learning Seminar

Title: Better than the best? Blending models and approaches in density-based clustering

Speaker: Alessandro Casa (University College Dublin)

Date: Mon 3rd February 2020 at 12:00PM

Location: Seminar Room SCN 1.25

Abstract: With the recent growth in data availability and complexity, and the associated outburst of elaborate modeling approaches, model selection tools have become a lifeline, providing objective criteria to deal with this increasingly challenging landscape. In fact, basing predictions and inference on a single model may be limiting if not harmful; ensemble approaches, which combine different models, have been proposed to overcome the selection step, and proven fruitful especially in the supervised learning framework. Conversely, these approaches have been scantily explored in the unsupervised setting. In this work we focus on the model-based clustering formulation, where a plethora of mixture models, with different number of components and parametrizations, is typically estimated. We propose an ensemble clustering approach that circumvents the single best model paradigm, while improving stability and robustness of the partitions. A new density estimator, being a convex linear combination of the density estimates in the ensemble, is introduced and exploited for group assignment. As opposed to the standard case, where clusters are associated to the components of the selected mixture model, we define partitions by borrowing the modal, or nonparametric, formulation of the clustering problem, where groups are linked with high-density regions. Staying in the density-based realm we thus show how blending together parametric and nonparametric approaches may be beneficial

from a clustering perspective.