



Working Group on Statistical Learning Seminar

Title: Parsimonious Bayesian Factor Analysis for modelling latent structures in spectroscopy data

Speaker: Alessandro Casa (University College Dublin)

Date: Mon 8th February 2021 at 12:00PM

Location: Online

Abstract: In recent years, within the dairy sector, animal diet and management practices have been receiving increased attention, in particular examining the impact of pasture-based feeding strategies on the composition and quality of milk and dairy products, in line with the increased prevalence of premium grass-fed dairy products appearing on market shelves. To date, there are limited testing methods available for the verification of grass-fed dairy and as a consequence these products are susceptible to food fraud and adulteration. Therefore, with this in mind, enhanced statistical tools studying potential differences among milk samples coming from animals on different feeding systems are required, thus providing increased security around the authenticity of the products. Infrared spectroscopy techniques are widely used to collect data on milk samples and to predict milk related traits and characteristics. While these data are routinely used to predict the composition of the macro components of milk, each spectrum also provides a reservoir of unharnessed information about the sample. The accumulation and subsequent interpretation of these data present a number of challenges due to their high-dimensionality and the relationships amongst the spectral variables. In this work, directly motivated by a dairy application, we propose a modification of the standard factor analysis to induce a parsimonious summary of spectroscopic data. The proposed procedure maps the observations into a low-

dimensional latent space while simultaneously clustering observed variables. The proposed method indicates possible redundancies in the data and it helps disentangle the complex relationships among the wavelengths. A flexible Bayesian estimation procedure is proposed for model fitting, providing reasonable values for the number of latent factors and clusters. The method is applied on milk mid-infrared (MIR) spectroscopy data from dairy cows on distinctly different pasture and non-pasture based diets, providing accurate modelling of the data correlation, the clustering of variables in the data and information on differences between milk samples from cows on different diets. <https://arxiv.org/pdf/2101.12499.pdf>

Join Zoom Meeting: <https://ucd-ie.zoom.us/j/68316324831>