Title: A probabilistic take on the partial least squares regression method

Dr Szymon Urbas

Abstract:

Partial least squares (PLS) regression is the go-to method for univariate and multivariate regression tasks in chemometrics, where the data are often high-dimensional and exhibit multicollinearity. PLS relies on projecting the data onto a lower-dimensional latent space; it is easy to use and provides accurate point predictions. However, a fundamental limitation of PLS is that it does not arise from the statistical inference of a probabilistic model. Tasks beyond basic point predictions require various ad hoc and post hoc modifications which are not always generalisable. We introduce the Bayesian partial least squares (BPLS) regression model with the of aim emulating the desirable properties of PLS whilst allowing for extensions based on the probabilistic formulation: (i) principled quantification of prediction uncertainty, (ii) eschewing the need to specify the latent dimension via Bayesian nonparametrics, and (iii) the potentials for extensions to data settings where samples are strictly dependent. I will illustrate BPLS through spectral data problems encountered in the dairy industry. This is joint work with Claire Gormley (UCD), Donagh Berry (Teagasc), Pierre Lovera, Rob Daly, and Alan O'Riordan (Tyndall Institute).