Title: Sliced Inverse Regression for skew data (**Research in Data Science**)

Advisors:

Brisbane: Geoffrey Mclachlan (g.mclachlan@uq.edu.au),
Sharon Lee (s.lee11@uq.edu.au) and
Hien Nguyen (h.nguyen5@latrobe.edu.au)

Local contact: Florence Forbes (florence.forbes@inria.fr) and
Stéphane Girard (stephane.girard@inria.fr)

Laboratories: Department of Mathematics, University of Queensland, Brisbane, Australia
Inria, LJK, Mistis team, Grenoble

Location: the research will ideally take place in part or totality in Brisbane under the co-supervision of the Inria Mistis team in Grenoble. Specific arrangements can be negociated with the applicant.

Project description:

Sliced Inverse Regression (SIR) [4] has been extensively used to reduce the dimension of the predictor space before performing regression. SIR is originally a model free method but it has been shown to actually correspond to the maximum likelihood of an inverse regression model with Gaussian errors [2]. Starting from this inverse regression formulation, authors in [1] have shown that Student distributed errors could be considered instead, leading to a so-called Student SIR method with improved robustness to outliers. The work in [1] makes use of the fact that using the Gaussian scale mixture formulation of the Student distribution, the inverse regression in SIR remains tractable via an Expectation-Maximization (EM) algorithm.
In practice, another useful extension of SIR would be to take into account the possibility that the data are coming from non-symmetric distributions. We propose to start again from the inverse regression formulation of SIR and consider errors with skew distributions. More specifically we would like to investigate the possibility to use Gaussian location and scale mixtures as proposed in [6] or the models in [5], as they provide skew and tractable distributions.
Then a second point is that even if the inverse regression model maximum likelihood remains tractable, there is no guarantee that the maximum likelihood solution is indeed the seek central subspace. In [2] a number of cases are studied besides the Gaussian case and a proof is given in the case of errors belonging to the exponential family. A first direction is then to start with skew distributions in the exponential family. Then, other skew distributions of interest may not belong to this family and we propose to investigate the so-called t-exponential family [3] as a possible alternative, which includes the Student distribution. More generally, this subject can be followed by a PhD.

Bibliography:

[1] Chiancone, A., Forbes, F., and Girard, S. (2016). Student sliced inverse regression. Computational Statistics & Data Analysis.
[2] Cook, R. D. (2007). Fisher lecture: Dimension reduction in regression. Statistical Science,22(1).
[3] Ding, N. and Wishwanathan, S. (2010). t-logistic regression. In Advances in Neural Information Processing Systems 23 (NIPS 2010).
[4] Li, K.-C. (1991). Sliced inverse regression for dimension reduction. Journal of the American Statistical Association, 86(414)..
[5] Lee, S.X. and McLachlan, G.J. Finite mixtures of canonical fundamental skew t-distributions: the unification of the restricted and unrestricted skew t-mixture models. Statistics and Computing 26, 573-589, 2016.
[6] Wraith, D. and Forbes, F. (2015). Location and scale mixtures of gaussians with exible tail behaviour: Properties, inference and application to multivariate clustering. Computational Statistics & Data Analysis.