LANDER Activity Report
2019

Title: Latent Analysis, Adversarial Networks, and DimEnsionality Reduction

Associate Team acronym: LANDER

Principal investigator (Inria): Florence Forbes, Mistis Inria Grenoble Rhône-Alpes, France

Principal investigator (Main team): Hien Nguyen, La Trobe University, Melbourne (Bundoora), Australia

Other participants:

- Queensland University of Technology, Brisbane, Australia
- University of Queensland, Brisbane, Australia.
- Swinburne University of Technology, Melbourne, Australia
- University of Caen, France

1 Hightlights

The main developments concern three topics: 1) mixtures of expert models which are a family of neural networks , 2) Bayesian inference based on the Approximate Bayesian Computation (ABC) method and 3) classification and clustering through a number of results regarding mixtures. More specifically concerning point 1) the collaboration between F. Forbes (Inria), H. Nguyen (Univ. La Trobe, Melbourne) and F. Chamroukhi (Univ. Caen) gave rise to a publication in which theoretical results on mixtures of experts in the multivariate case are detailed [1]. For point 2), work involving the Inria team (J. Arbel, F. Forbes, and the post-doctoral fellow H. Lu) and La Trobe University (H. Nguyen) has also been finalised and submitted to a journal [5]. This work concerns the use of original metrics in the ABC methods known as important sampling ABC and will be presented at a Conference in Munich in 2020, UQ2020. Point 3) covers in a broad sense various topics involving mixture models [3, 8]. Other ongoing research relates to other types of approximations by mixtures with guaranteed properties ([7] submitted) and to the development of efficient clustering algorithms for big data ([6] under revision), and in the non-Gaussian case [2].

2 Scientific achievements

2.1 Approximation results regarding the multiple-output Gaussian gated mixture of linear experts model

Participants: H. Nguyen, F. Chamroukhi, F. Forbes

Mixture of experts (MoE) models are a class of artificial neural networks that can be used for functional approximation and probabilistic modeling. An important class of MoE models is the class of mixture of linear experts (MoLE) models, where the expert functions map to real topological output spaces. Recently, Gaussian gated MoLE models have become popular in applied research. There are a number of powerful approximation results regarding Gaussian gated MoLE models, when the output space is univariate. These results guarantee the ability of Gaussian gated MoLE mean functions to approximate arbitrary continuous functions, and Gaussian gated MoLE models themselves to approximate arbitrary conditional probability density functions. We utilized and extended upon the univariate approximation results in order to prove useful results for situations where the output spaces are multivariate. We did this by proving lemmas regarding the combination of univariate MoLE models, which are interesting in their own rights [1].

2.2 Approximate Bayesian computation via the energy statistic

Participants: H. Nguyen, J. Arbel, H. Lu, F. Forbes

Approximate Bayesian computation (ABC) has become an essential part of the Bayesian toolbox for addressing problems in which the likelihood is prohibitively expensive or entirely unknown, making it intractable. ABC defines a quasi-posterior by comparing observed data with simulated data, traditionally based on some summary statistics, the elicitation of which is regarded as a key difficulty. In recent years, a number of data discrepancy measures bypassing the construction of summary statistics have been proposed, including the Kullback–Leibler divergence, the Wasserstein distance and maximum mean discrepancies. Here we propose a novel importance-sampling (IS) ABC algorithm relying on the so-called *two-sample energy statistic*. We establish a new asymptotic result for the case where both the observed sample size and the simulated data sample size increase to infinity, which highlights to what extent the data discrepancy measure impacts the asymptotic pseudo-posterior. The result holds in the broad setting of IS-ABC methodologies, thus generalizing previous results that have been established only for rejection ABC algorithms. Furthermore, we propose a consistent V-statistic estimator of the energy statistic, under which we show that the large sample result holds. Our proposed energy statistic based ABC algorithm is demonstrated on a variety of models, including a Gaussian mixture, a moving-average model of order two, a bivariate beta and a multivariate g-and-k distribution. We find that our proposed method compares well with alternative discrepancy measures [5].

2.3 Model based clustering and classification of functional data

Participants: H. Nguyen, F. Chamroukhi

Complex data analysis is a central topic of modern statistics and learning systems which is becoming of broader interest with the increasing prevalence of high dimensional data. The challenge is to develop statistical models and autonomous algorithms that are able to discern knowledge from raw data, which can be achieved through clustering techniques, or to make predictions of future data via classification techniques. Latent data models, including mixture model based approaches, are among the most popular and successful approaches in both supervised and unsupervised learning. Although being traditional tools in multivariate analysis, they are growing in popularity when considered in the framework of functional data analysis (FDA). FDA is the data analysis paradigm in which each datum is a function, rather than a real vector. In many areas of application, including signal and image processing, functional imaging, bioinformatics, etc., the analyzed data are indeed often available in the form of discretized values of functions, curves, or surfaces. This functional aspect of the data adds additional difficulties when compared to classical multivariate data analysis. We reviewed approaches for model based clustering and classification of functional data. We considered well grounded statistical models along with efficient algorithmic tools to address problems regarding the clustering and the classification of these functional data, including their heterogeneity, missing information, and dynamical hidden structures. The presented models and algorithms were illustrated via real world functional data analysis problems from several areas of application [3].

2.4 Finite mixtures of multiple scaled Generalized Hyperbolic distributions using a Bayesian approach

Participants: D. Wraith, F. Forbes

The multiple scaled Generalized Hyperbolic (MSGH) family is an attractive and flexible family of probability distributions, which can provide different degrees of heavy-tailedness and asymmetric properties for each dimension of the variable space. This family contains some special cases and limiting distributions, such as the Gaussian, Student-t, Normal Inverse Gaussian (NIG), variance-gamma distributions and subsequent multiple scaled versions of all of these. This paper presents a Bayesian approach to estimation of the multiple scaled Generalized Hyperbolic (MSGH) family. In particular, we develop an MCMC approach using Gibbs sampling which has some advantages in this context as most of the Gibbs sampling updates are available in closed form. The approach is illustrated, and the performance examined, by applying the MSGH to mixture model problems using simulated and real data which present challenging clustering problems [8].

2.5 Mini-batch learning of exponential family finite mixture models

Participants: H. Nguyen, G. McLachlan, F. Forbes

Mini-batch algorithms have become increasingly popular due to the requirement for solving optimization problems, based on large-scale data sets. Using an existing online expectation-maximization (EM) algorithm framework, we demonstrate how mini-batch (MB) algorithms may be constructed, and propose a scheme for the stochastic stabilization of the constructed mini-batch algorithms. Theoretical results regarding the convergence of the mini-batch EM algorithms are presented. We then demonstrate how the mini-batch framework may be applied to conduct maximum likelihood (ML) estimation of mixtures of exponential family distributions, with emphasis on ML estimation for mixtures of normal distributions. Via a simulation study, we demonstrate that the mini-batch algorithm for mixtures of normal distributions can outperform the standard EM algorithm. Further evidence of the performance of the mini-batch framework is provided via an application to the famous MNIST data set [6].

2.6 Approximation by finite mixtures of continuous density functions that vanish at infinity

Participants: T. Nguyen, H. Nguyen, F. Chamroukhi

Given sufficiently many components, it is often cited that finite mixture models can approximate any other probability density function (pdf) to an arbitrary degree of accuracy. Unfortunately, the nature of this approximation result is often left unclear. We prove that finite mixture models constructed from pdfs in C_0 can be used to conduct approximation of various classes of approximands in a number of different modes. That is, we prove approximands in C_0 can be uniformly approximated, approximands in C_b can be uniformly approximated on compact sets, and approximands in $\mathcal{L}_p$ can be approximated with respect to the $\mathcal{L}_p$, for $p \in [1, \infty)$. Furthermore, we also prove that measurable functions can be approximated, almost everywhere [7].

2.7 On strict sub-Gaussianity, optimal proxy variance and symmetry for bounded random variables

Participants: J. Arbel, H. Nguyen

We investigated the sub-Gaussian property for almost surely bounded random variables. If sub-Gaussianity per se is de facto ensured by the bounded support of said random variables, then exciting research avenues remain open. Among these questions is how to characterize the optimal sub-Gaussian proxy variance? Another question is how to characterize strict sub-Gaussianity, defined by a proxy variance equal to the (standard)variance? We addressed the questions in proposing conditions based on the study of functions variations. A particular focus was given to the relationship between strict sub-Gaussianity and symmetry of the distribution. In particular, we demonstrated that symmetry was neither sufficient nor necessary for strict sub-Gaussianity. In contrast, simple necessary conditions on the one hand, and simple sufficient conditions on the other hand, for strict sub-Gaussianity are provided. These results were illustrated via various applications to a number of bounded random variables, including Bernoulli, beta, binomial, uniform, Kumaraswamy, and triangular distributions [2].

3 Future work and Planned activities

The next year will be devoted to the finalization of the mentioned working papers, with further developments regarding ABC and clustering algorithms for big data. Additional work that has just been initiated relates to the development of majorization-minimization algorithms for the computation of empirical and theoretical geometric multivariate expectiles. We also plan to start investigating extensions of inverse regression techniques for asymmetric (skew) data and selection criteria that better account for potential heavy tails.

3.1 Majorization-minimization algorithms for the computation of empirical and theoretical geometric multivariate expectiles

Participants: H. Nguyen, A. Usseglio-Carlève, S. Girard

One of the most popular risk measures is the Value-at-Risk (VaR) introduced in the 1990's. The Value-at-Risk however suffers from several weaknesses. A possible coherent alternative risk measure is based on expectiles. Compared to quantiles, the family of expectiles is based on squared rather than absolute error loss minimization. The flexibility and virtues of these least squares analogues of quantiles are now well established in actuarial science, econometrics and statistical finance. Both quantiles and expectiles were embedded in the more general class of M-quantiles as the minimizers of a generic asymmetric convex loss function. It has been proved very recently that the only M-quantiles that are coherent risk measures are the expectiles. In this ongoing work we address the practical computation of such expectiles using majorization-minimization algorithms.

3.2 Sliced Inverse Regression for skew data

Participants: S. Lee, G. McLachlan, F. Forbes, S. Girard

Sliced Inverse Regression (SIR) [13] 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 [11]. Starting from this inverse regression formulation, we have shown [10] that Student distributed errors could be considered instead, leading to a so-called Student SIR method with improved robustness to outliers. The work in [10] 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 [16] or the models in [14], 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 [11] 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 [12] as a possible alternative, which includes the Student distribution.

3.3 Model selection criteria for data with heavy tails.

Participants: D. Wraith, F. Forbes

The interest is in fitting well in the tails/extremes. This would require to extend traditional selection criteria. We will first investigate the possibility to design weighted versions of the BIC or the log-likelihood where the weights represent the importance of the tails. This relates to previous work on the multiple scaled distributions [15, 16] where tailweight is allowed to vary across dimensions. The use of the weights is quite similar to the original idea of using the weights to explore better some regions/outliers but in this case the weights would be rather used to assess goodness of fit. Priors could be used on the weights for each observation as a form of penalty.

4 Visits, Events and Network

4.1 Visits

- F. Forbes (Inria) spent 2 weeks at La Trobe university, Melbourne in July 2019.
- F. Chamroukhi (U. Caen) visited UQ in Brisbane for 2 months (July and August) and spent 1 week at La Trobe in July 2019.
- A. Usseglio-Carleve (Inria) spent 2 weeks at La Trobe in Melbourne in October 2019.
- N. Karavarsamis (La Trobe) will spend 1 week at Inria Grenoble in November 2019
- H. D. Nguyen (La Trobe) will to spend 2 weeks at Inria Grenoble in November 2019.
- D. Wraith (QUT) will spend 2 weeks at Inria Grenoble in December 2019 and January 2020.
- K-A. Le Cao (Melbourne University) will spend one week in Marseille for a co-organized Winter School at CIRM and 2 days at Inria Grenoble in March 2020.

4.2 Organized events

- Research School on Statistics and Data Science 2019, RSSDS2019, La Trobe University, Melbourne, 24-26 July
- Statistics Seminar Lab Jean Kuntzmann Grenoble: Natalie Karavarsamis from La Trobe University, Melbourne. Estimating occupancy and the two-stage approach, November 14, 2019, Grenoble, France.
- Kim-Anh Le Cao's lecture at CIRM Winter school on Networks and molecular biology, 2-6 March 2020, Marseille, France.
- Seminar Inria Grenoble Rhone-Alpes, Kim-Anh Le Cao from Melbourne University, March 9-10, 2020, Grenoble, France.

The different visits made it possible to develop the scientific work described above but also to make other contacts in order to pursue collaborations and exchanges between researchers (Melbourne, Brisbane on the Australian side, Grenoble and Caen on the French side).

We established a first contact between the Grenoble Data Institute (Data@UGA project) and two similar institutes in Melbourne:

- Swinburn Data Science Research Institute, meetings with Kai Qin and Timos Sellins who is coordinating this project.
- The Trobe Business School Research Center for Data Analytics and Cognition, meetings with Kok-Leong Ong and with Wenny Rahayu (Head of School of Engineering and Mathematical sciences at La Trobe) to discuss concrete modalities for student exchange,

In addition, several team members have participated to the "Research School on Statistics and Data Science" workshop held in Melbourne in July 2019

5 Publications from the team

5.1 Journal publications

[1] Hien D. Nguyen, Faicel Chamroukhi, Florence Forbes. Approximation results regarding the multiple-output Gaussian gated mixture of linear experts model. Neurocomputing, Elsevier, 2019.

[2] Julyan Arbel, Olivier Marchal, and Hien D. Nguyen. On strict sub-Gaussianity, optimal proxy variance and symmetry for bounded random variables. ESAIM: Probability & Statistics, forthcoming, 2019.

[3] Faicel Chamroukhi, Hien D. Nguyen. Model-based clustering and classification of functional data. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 9. 2019.

5.2 Conference publication

[4] Hien D. Nguyen, Julyan Arbel, Hongliang Lu, and Florence Forbes. Approximate Bayesian computation via the energy statistic, UQ2020, Munich

5.3 Working papers

[5] Hien D. Nguyen, Julyan Arbel, Hongliang Lu, and Florence Forbes. Approximate Bayesian computation via the energy statistic.

[6] Hien D. Nguyen, Florence Forbes and Geoffrey J. McLachlan. Mini-batch learning of exponential family finite mixture models.

[7] Tin Nguyen, Hien D. Nguyen, Faicel Chamroukhi, Geoffrey J. McLachlan (2019). Approximation by finite mixtures of continuous density functions that vanish at infinity.

[8] Mohsen Maleki, Darren Wraith and Florence Forbes. Finite mixtures of multiple scaled Generalized Hyperbolic distributions using a Bayesian approach.

[9] Antoine Usseglio-Carleve, Hien D. Nguyen, Stéphane Girard. Development of majorization-minimization algorithms for the computation of empirical and theoretical geometric multivariate expectiles.

5.4 References

[10] Chiancone, A., Forbes, F., and Girard, S. (2016). Student sliced inverse regression. Computational Statistics & Data Analysis.

[11] Cook, R. D. (2007). Fisher lecture: Dimension reduction in regression. Statistical Science,22(1).

[12] Ding, N. and Wishwanathan, S. (2010). t-logistic regression. In Advances in Neural Information Processing Systems 23 (NIPS 2010).

[13] Li, K.-C. (1991). Sliced inverse regression for dimension reduction. Journal of the American Statistical Association, 86(414)..

[14] 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.

[15] F. Forbes and D. Wraith. A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. *Statistics & Computing*, 24(6):971–984, 2014.

[16] D. Wraith and F. Forbes. Location and scale mixtures of Gaussians with flexible tail behaviour: Properties, inference and application to multivariate clustering. *Computational Statistics & Data Analysis*, 90:61–73, 2015.