Large dimensional analysis of LS-SVM transfer learning: application to PolSAR classification

1. Introduction

Interpretation and automatic recognition of satellite images (radar) are major challenges for environmental monitoring. Remote sensing is a tool for understanding and responding to environmental issues and current ecological systems. Although the methods derived from the theory of machine learning have provided major advances in the optical recognition of 'images, these algorithms cannot simply be transferred to radar imaging applications. The backscattering properties of materials depend on the frequency of the incident wave. In some cases, the size of the data analyzed makes automatic analysis difficult. In modern imaging methods using automatic learning, large-dimensional data alone embodies a highly active area of image processing, and there is a need for attention to new methods of imaging. To designate the difficulty of analyzing the performance of classical algorithms in large dimensions, the term “dimensionality curse” is sometimes used.

It is necessary to understand new accurate and robust classifiers to automatically recognize objects and models in radar imagery, more precisely synthetic aperture radar polarimetric images (PolSAR). The polarimetric SAR system provides rich information on target scattering by using different polarizations to alternately transmit and receive radar signals, which has become an important tool for ground exploration. the analysis of supervised and unsupervised learning supervised using the theory of random matrices has given rise to a rich corpus. These methods allow us to better understand the classification performances of the different statistical models which best characterize the SAR data. Contemporary studies in the field of large statistical dimensions shed new light on the performance of machine learning methods such as [3], [14], [16] and [21].

However, insufficient data training available is an unavoidable problem in the field of satellite imagery. Data collection is complex and expensive, which makes it extremely difficult to create a large-scale, high-quality annotated data set. Moreover, even though we get a training dataset by paying a high price, it is very easily obsolete and therefore cannot be effectively applied in new tasks. one can be interested in two structurally different approaches to a learning problem for which there is little label on the data, which is often the case in radar data: learning not supervised or transfer learning (respectively more commonly referred to as clustering and transfer learning).

2. Context

The aim of this work is to develop new learning methods, supervised or not, precise, robust and efficient in terms of computation, allowing to automatically classify or recognize models, objects or classes in radar imaging applications. The main novelty of the proposed approach is to exploit the recent tools from the study of random matrices. These methods allow in particular, for the first time, to obtain a theoretical understanding of classification performance for different data models and thus provide guarantees and performance limits. These theoretical performances of classification algorithms are one of the missing pillars data-driven methods, which are available here through mathematical concentration arguments for models verifying large-scale hypotheses.

An interesting way to understand large-scale framework and particularly to observe the so-called "curse of dimensionnality" is to observe the lack of relevance of the Euclidean distance, in a regime of large dimension. In the theory of random matrices, we call "regime n, p", the regime in which each data sample has a dimension which is of the same order than the number of samples. This regime allows an asymptotical analysis.

We observe a phenomenon of concentration of the distance between two draws of random vector variables around a deterministic value noted τ. This constant is fixed as a function of the parameters of the distributions of the data and therefore has an estimate which is accessible to us.

Let a database $X=[x_1,\ldots,x_n]\in R^{p\times n}$ and x_i , x_j be the elements of R^p from two different Gaussian mixtures in the sample *X*, respectively $N(\mu_1,C_1)$ and $N(\mu_2,C_2)$. Each Gaussian mixture denoting a class. Under certain assumptions, in particular making the distinction between two different classes non-trivial, we have uniformly, for *n* and *p* at the same scale, that is to say pending jointly towards infinity:

$$\frac{1}{p}\|x_i-x_j\|\rightarrow\tau\, a.s. when\, n,p \quad \infty$$

This phenomenon of concentration partly explains the practical difficulties of classical algorithms in large dimensions. At first glance, this would mean that regardless of the prior difference between the distributions, their Euclidean distance will tend towards the same value.

Subsequently, we will see that large-dimensional analysis allows not only to understand the curse of dimensionality but to benefit from it via developments around nonlinear (but continuously derivable) functions.

3. References

[3] R. Couillet and F. Benaych-Georges. Kernel spectral clustering of large dimensional data. Electronic Journal of Statistics, page 1393–1454, 2016.
[4] R. Couillet and A. Kammoun. Random matrix improved subspace clustering. In 2016 50thAsilomar Conference on Signals, Systems and Computers, pages 90–94. IEEE, 2016.
[6] M. Harandi et al. Joint dimensionality reduction and metric learning: A geometric take. Volume 70,2017.
[7] N. El Karoui et al. The spectrum of kernel random matrices.Annals of statistics, 38(1):1–50,2010.
[8] P. Formont et al. Statistical classification for heterogeneous polarimetric sar images.IEEE Journal of Selected Topics in SignalProcessing, 5(3):567–576, 2011.
[10] S. Xu et al. Multi-task least-squares support vector machines.Multimedia tools and applications,71(2):699–715, 2014.
[11] T. Evgeniou and M. Pontil. Regularized multi–task learning.Proceedings of the tenth ACM-SIGKDD international conference on Knowledge discoveryand data mining., pages 109–117, 2004.
[12] Z. Fan and A. Montanari. The spectral norm of random inner-product kernel matrices. ProbabilityTheory and Related Fields, 173(1):27–85, 2019.
[14] Z. Liao and R. Couillet. A large dimensional analysis of least squares support vector machines.IEEE Transactions on Signal Processing, 67(4):1065–1074, 2019.
[15] U. V. Luxburg. A tutorial on spectral clustering.Statistics and computing, 17(4):395–416, 2007.
[16] X. Mai and R. Couillet. A random matrix analysis and improvement of semi-supervised learningfor large dimensional data.The Journal of Machine Learning Research, 19(1):3074–3100, 2018.
[18] S. J. Pan and Q. Yang. A survey on transfer learning.IEEE Transactions on Knowledge and Data Engineering, 22(10):1345–1359, 2010.
[20] J. A. K. Suykens and J. Vandewalle. Least squares support vector machine classifiers.Neuralprocessing letters, vol. 9, no. 3, page 293–300, 1999.
[21] M. Tiomoko and R. Couillet. Large dimensional analysis and improvement of multi task learning.Journal of Machine Learning Research, 2020.
[22] M. Tiomoko, C. Louart, and R. Couillet. Large dimensional asymptotics of multi-task learning.2020.
[23] V. Vapnik. The nature of statistical learning theory.Springer science and business media, 2013.
[24] C. Xiuyuan and S. Amit. The spectrum of random inner-product kernel matrices.RandomMatrices: Theory and Applications, 2(04):1350010, 2013.
[25] F. Zhuang, Z. Qi, K. Duan, D. Xi, and Y. Zhu. A comprehensive survey on transfer learning.Proceedings of the IEEE, pages 43–76, 2020.