MASCOT-NUM 2022
Clermont Ferrand, France
7-9 June 2022

Auto-Associative Models, a generalized PCA

Valentin Pibernus
Phimeca Engineering

Supervisor(s): Stéphane Girard (INRIA Grenoble), Serge Iovleff (UTBM) and Sylvain Girard (Phimeca Engineering)

PhD expected duration: Oct. 2019 - Sep. 2022

Address: Phimeca Engineering, 18 Bd de Reuilly, 75011, Paris

E-mail: pibernus@phimeca.com

Abstract:

The manipulation of high-dimensional data is a problem that has required the creation of tools of increasing complexity in recent years, notably neural networks, deep networks and other black-box models. In this presentation, the class of problems we are interested in is analogous to constructing a map between a set G of high dimension p and a set S of low dimension d. To solve this problem, dimension reduction methods are used. These methods aim to capture the information carried by the data in G and to transcribe it into a low-dimensional space, using a reduced number of variables. These methods are known to be efficient when the intrinsic dimension of the data is lower than the dimension of G, for example when our data set is a d-manifold of topological dimension d contained into a space of dimension p.

In this talk, we propose Auto-Associative Models (AAM) as a non-linear generalization of Principal Component Analysis (PCA). Auto-Associative Models are a self-supervised dimension reduction method. They are used to approximate a data set by a manifold with each iteration increasing the dimension of the manifold by one. Similar to PCA, the AAM algorithm iteratively searches for a projection direction maximizing a given criterion and then builds at a recovery function estimating the initial coordinates from the projected coordinates.

This innovative method proposed by Stéphane Girard in [2] is based on two major ideas:

1. the search for an 'invertible' projection direction, such a projection would allows the reconstruction of the initial data set from the projected coordinates. The criterion proposed by the AAM is a topological criterion. We try to preserve neighborhood relations between points. This idea allows to reveal more complex structures of the space by 'unfolding' the manifold.
2. the construction of a non-linear recovery function, which estimates the priors of the projected points. It is called the inverse of the projection or recovery function.

Taken together, these two ideas enable to reconstruct very high dimensional data sets whose intrinsic dimension is low. An example might be the translated or spread Gaussian functions used in [1]. The problem consists of the reconstruction of a set of Gaussian functions. These functions are represented in a space of dimension n by discretizing the set of definition of the functions. They are characterized by their mean and their variance. With fixed variance (respectively fixed mean), they belong to a set of intrinsic dimension $k = 1$ contained in $\mathbb{R}^n$. We are able to reconstruct these Gaussian functions in one iteration of the AAM. In other words, where PCA fails to reconstruct a subspace of sufficiently small dimension, the AAM method manages to reconstruct a subspace whose dimension corresponds to the intrinsic dimension of the problem.

However, some problems resist this method. If we consider the previous example without fixing one of the two parameters, we study a manifold of dimension 2 contained in a space of dimension n. This set is not reconstructable by the method of Auto-Associative Models. In this talk, we will introduce the AAM and then we will use these examples to expose the limits of the AAM and to introduce an extension of the method to solve a wider class of problems.

The proposed extension is inspired by the approach used in Sliced Inverse Regression (SIR) and uses a second data set to supervise the AAM.

References

[1] Keinosuke Fukunaga and David R. Olsen. An algorithm for finding intrinsic dimensionality of data. *IEEE Transactions on Computers*, C-20(2):176–183, feb 1971.

[2] Stéphane Girard and Serge Iovleff. Auto-associative models, nonlinear principal component analysis, manifolds and projection pursuit. pages 202–218, 2008.

Short biography – I am a graduate engineer from the Ecole Centrale de Nantes. I am working on auto-associative models in the framework of a collaboration contract between Phimeca and INRIA. This contract is funded by Phimeca and co-financed by AMIES