Neurocomputing 162 (2015) 209–217

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

An identifying function approach for determining parameter structure of statistical learning machines

Zhi-Yong Ran [a,*], Bao-Gang Hu [b]

[a] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[b] NLPR&LIAMA, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

ARTICLE INFO

Article history:
Received 2 September 2014
Received in revised form
12 March 2015
Accepted 16 March 2015
Communicated by S. Fiori
Available online 2 April 2015

Keywords:
Identifying function
Structural identifiability
Statistical learning machine
Kullback–Leibler divergence
Parameter redundancy
Reparameterization

ABSTRACT

This paper presents an *identifying function* (*IF*) approach for determining parameter structure of *statistical learning machines* (*SLMs*). This involves studying three related aspects: *structural identifiability* (*SI*), *parameter redundancy* (*PR*) and *reparameterization*. Firstly, by employing the Rank Theorem in Riemann geometry, we derive an efficient identifiability criterion by calculating the rank of the derivative matrix (DM) of IF. Secondly, we extend the previous concept of IF to *local IF* (*LIF*) for examining local parameter structure of SLMs, and prove that the Kullback–Leibler divergence (KLD) is such a proper LIF, thus relating the LIF approach to several existing criteria. Lastly, an analytical approach for solving *minimal reparameterization* in parameter-redundant models is established. The dimensionality of the minimal reparameterization can be used to characterize the *intrinsic parameter dimensionality* of model. We compare the IF approach with existing criteria and discuss its pros/cons from theoretical and application viewpoints. Several model examples from the literature are presented to study their parameter structure.

1. Introduction

Learning machines such as neural networks and support vector machines have been widely applied to a variety of scientific areas. As [1–3] show, it is useful to describe learning machines as a family of probability density functions (PDFs). For example, from the statistical viewpoint, the least squared learning in neural networks can be interpreted as nonlinear regression which approximates the conditional expectation of the output given an input, and the resulting least squared estimator is equal to the maximum likelihood estimator [1,3]. While in support vector machines for classification, the exponential of negative hinge loss can be interpreted as a Gaussian scale mixture; this in turn opens the door to Bayesian methods for setting the hyper-parameters of the original model parameters [4], to just name a few. For a detailed statistical treatment in machine learning models, one can see [1–3] and the references therein. The statistical perspective is clearly an effective one which provides deep insight into machine learning methods. For stressing on the statistical modeling nature, in this study, we call such models *statistical learning machines* (*SLMs*).

Structural identifiability (*SI*) is concerned with theoretical uniqueness of model parameters determined from an underlying statistical family [5–7]. In a general sense the SI is just one aspect of a larger problem, the inverse problem [8], which basically encompasses SI and identification (e.g., objective function, regularization, and learning algorithm). For a rigorous treatment of SI problem, one should distinguish between the concepts of global identifiability and local identifiability. Roughly speaking, an SLM is said to be *globally identifiable* if different parameter values lead to different PDFs throughout the parameter space; it is said to be *locally identifiable* if there exists a small distance such that any two parameters giving the same PDF must be separated at least by that distance [5–7]. *Parameter redundancy* (*PR*) occurs if the model can be rewritten in terms of a smaller set of parameters [9]. The concept that intimately relates to SI and PR is *parameter dependence* (*PD*) in the sense that a certain subvector of parameter can be expressed as the function of the remaining one [10].

Besides being an important way to enhance model transparency and comprehensibility [11,12], the SI is also a necessary prerequisite for system modeling and parameter estimation [13]. Typically, if an SLM has hierarchial structures [17,18], latent variables [1,16], state variables [13], nuisance parameters [19] or coupled submodels [11,12], the model may be unidentifiable. Due to the universal existence of nonidentifiability, Watanabe pointed out that “almost all learning machines are singular” [18]. Moreover, the SI issue has a

* Corresponding author. Tel.: +86 13650593660.
E-mail addresses: ranzy@cqupt.edu.cn (Z.-Y. Ran), hubg@nlpr.ia.ac.cn (B.-G. Hu).

http://dx.doi.org/10.1016/j.neucom.2015.03.050
0925-2312/

close connection with a range of subjects such as variational Bayesian matrix factorization [14], low-rank matrix completion [15], latent factor model [16], and probabilistic PCA [1]. The utility and importance of SI study can be further recognized in, e.g., physically interpretable (sub-) models [11–13,20], singular learning theory [18,21,22], statistical inference [5,17,22], and learning algorithm and learning dynamics [17,23]. Therefore, it is of special importance to study SI in the field of machine learning.

In this paper, we report an extension of existing investigations on determining parameter structure of SLMs in the following two directions:

- *SI analysis*: In our previous studies, Yang et al. [11,12] considered the SI problem in *generalized-constraint neural network* model, and derived identifiability theorems for *Single-input Single-output* (*SISO*) and *Multiple-input Single-output* (*MISO*) models. However, their theorems cannot deal with *Multiple-input Multiple-output* (*MIMO*) models. In [24], Ran et al. derived the identifiability result for MIMO models. For a detailed description of relevant results, one can see [11,12,24] and the references therein. Hence, this paper is a further investigation built upon [10–12,24,26], and we expect to link the present study to existing criteria, thus providing a deeper insight into SI theory.
- *PR and reparameterization*: We develop an analytical method for reparameterization when PR is detected, and show that the dimensionality of the resulting *minimal reparameterization* can be used to characterize the *intrinsic parameter dimensionality* of model. Compared with the method in [9] which is applicable for exponential family, the present method is workable in more generic statistical settings. Moreover, it processes reparameterization in a global way, while the method in [5] is merely a local one as it is based on examining local identifiability. Therefore, this paper is an extension of [5,9], and we further expect to study the PR problem from a wider range of models.

The goal of this paper is to present a systematic treatment for determining parameter structure of SLMs. To this end, we start by making use of the existing concept of *identifying function* (*IF*) [5] to give an efficient identifiability criterion with the help of the Rank Theorem in Riemann geometry [29]. The key idea is to transform the SI problem into the problem of examining the injectivity of the mapping IF. The derived method works by calculating the rank of the *derivative matrix* (*DM*) of IF. Then, we extend the definition of IF to *local identifying function* (*LIF*) for determining local parameter structure of SLMs, and prove that the *Kullback–Leibler divergence* (*KLD*) is such a proper LIF. We further demonstrate that several well-established criteria, such as Fisher Information Matrix (FIM) [7], Kullback–Leibler divergence equation (KLDE) [6], and local least squared (LLS) [27], can be directly derived from the LIF theme, thus revealing the common basis underpinning these identifiability criteria. Lastly, an analytical method for reparameterization by constructing *functionally independent* (see Definition 10) parametric functions is established. It is worth noting that the procedure can be implemented in a step-by-step manner.

The main contribution of this paper is given from the following two aspects:

1. Based on the IF theme, by employing the Rank Theorem in Riemann geometry, we derive an efficient identifiability criterion. Moreover, we extend the previous definition of IF to LIF for determining local parameter structure of SLMs, and prove that the KLD is such a proper LIF.
2. Based on the DM of IF, we develop an analytical method for reparameterization when PR is detected, and show that the dimensionality of the minimal reparameterization can be used to characterize the intrinsic parameter dimensionality of model.

The remainder of this paper is organized as follows. Section 2 introduces the basic concepts and reviews the existing methods. In Section 3, we present an IF approach for dealing with SI problem, and extend the previous definition of IF to LIF for studying local parameter structure of SLMs. In Section 4, based on the IF approach, we study the PR problem and present an analytical method for reparameterization. In Section 5, several model examples from the literature are presented to examine their parameter structure. Section 6 concludes this paper and points out the direction for future work.

2. SI analysis: basic concepts and existing methods

Let $p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ be the conditional PDF of an output $\mathbf{y}\in\mathbb{R}^m$ for a given input $\mathbf{x}\in\mathbb{R}^n$ and a parameter $\boldsymbol{\theta}\in\mathbb{R}^k$. Here $p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ is referred as a learning machine [22]. For highlighting the statistical modeling nature, the model in this study is called an SLM. The SI analysis for SLMs is concerned with theoretical uniqueness of model parameters determined from a family of underlying PDFs. Formally, the SI is defined in terms of the mapping $\boldsymbol{\theta}\mapsto p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ being one-to-one. Let $p(\mathbf{x})$ be the PDF of $\mathbf{x}$, which is positive for *almost everywhere* (*a.e.*) $\mathbf{x}\in\mathbb{R}^n$, $\mathbf{z}=(\mathbf{x},\mathbf{y})$ and $p(\mathbf{z},\boldsymbol{\theta})=p(\mathbf{x},\mathbf{y},\boldsymbol{\theta})=p(\mathbf{x})p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ be the joint PDF of $\mathbf{x}$ and $\mathbf{y}$, the SI can be equivalently defined as the mapping $\boldsymbol{\theta}\mapsto p(\mathbf{z},\boldsymbol{\theta})$ being one-to-one. For many application scenarios, the PDF $p(\mathbf{z},\boldsymbol{\theta})$ can be replaced by the characteristic function of $\mathbf{z}$ (the Fourier transformation of $p(\mathbf{z},\boldsymbol{\theta})$) since the correspondence between a PDF and its characteristic function is one-to-one [31]. In the present paper, a special emphasis is placed on *nonlinear* SLM which is nonlinear with respect to its unknown parameter vector. Hereafter, for notational simplicity, the SLM is denoted by the probability measure [31] $P_{\boldsymbol{\theta}}$ and will be interchangeably used with its PDF $p(\mathbf{z},\boldsymbol{\theta})$.

Following [5–7], we give the following definitions.

Definition 1. An SLM $P_{\boldsymbol{\theta}}$ is globally identifiable if $P_{\boldsymbol{\theta}_1}=P_{\boldsymbol{\theta}_2}\Rightarrow\boldsymbol{\theta}_1=\boldsymbol{\theta}_2,\forall\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in\mathbb{R}^k$. An SLM is locally identifiable if for every $\boldsymbol{\theta}\in\mathbb{R}^k$, there is an open neighborhood $N(\boldsymbol{\theta})$ of θ such that $P_{\boldsymbol{\theta}_1}=P_{\boldsymbol{\theta}_2}\Rightarrow\boldsymbol{\theta}_1=\boldsymbol{\theta}_2,\forall\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in N(\boldsymbol{\theta})$.

If a parameter point $\boldsymbol{\alpha}\in\mathbb{R}^k$ is of particular interest, for example, $\boldsymbol{\alpha}$ is assumed to be the real value (or critical value) of the model parameter, we give the following definition.

Definition 2. An SLM $P_{\boldsymbol{\theta}}$ is globally identifiable at α if $P_{\boldsymbol{\theta}}=P_{\boldsymbol{\alpha}}$, $\boldsymbol{\theta}\in\mathbb{R}^k\Rightarrow\boldsymbol{\theta}=\boldsymbol{\alpha}$. An SLM is locally identifiable at α if there is an open neighborhood $N(\boldsymbol{\alpha})$ of α such that $P_{\boldsymbol{\theta}}=P_{\boldsymbol{\alpha}},\boldsymbol{\theta}\in N(\boldsymbol{\alpha})\Rightarrow\boldsymbol{\theta}=\boldsymbol{\alpha}$.

As a matter of fact, the statements in Definition 1 and 2 can be relaxed to hold with an exception of a subset of zero measure in parameter space, since there may exist a manifold of atypical parameters which constitutes a zero-measure subset [33,34]. For example, consider the model $y=\theta_1 f(\theta_2 x)+\epsilon$, where f is a nonlinear function, the noise $\epsilon\sim\mathcal{N}(0,1)$ and parameter $\boldsymbol{\theta}=(\theta_1,\theta_2)\in\mathbb{R}^2$, all the parameters in the manifold $\{(0,\theta_2):\theta_2\in\mathbb{R}\}$ give rise to the same PDF. However, the model is globally identifiable in usual dominated situations because this manifold is a zero-measure subset in the parameter space. This type of singularity can also be exemplified by familiar learning machines, such as multi-layer perceptrons, Gaussian mixtures, and radical basis function networks. For more details, one can see [17] and the references therein.

Remark 1. It should be noted that the above relaxation does not mean that the singularity in a zero-measure subset has no influence at all in the study of machine learning. The importance of singularity should not be underestimated but be emphasized. For example, Watanabe et al. study the singular learning theory (e.g., training and generalization error, and Bayesian inference) by using algebraic geometry [18,21,22,35], whereas Amari et al. study the behavior of