Deep Learning and statistical model choice

Proposed supervisor: Petter Mostad
Research area (forskarutbildningsämne): Mathematical Statistics

Overview

A fundamental issue in both Artificial Intelligence (AI) and in mathematical statistics is model choice. In AI, the objective may be to find a specific neural network that yields good results in a particular context; in mathematical statistics, one may be selecting from a wider range of stochastic models, guided by the context.

Whereas many other methodological issues in statistics are resolved when using a strict Bayesian paradigm, the initial choice of model lies outside that paradigm by definition. In addition to the many practical procedures for model choice that are in use, a number of theoretical frameworks have been developed. Several frameworks are based on information criteria such as the Deviance Information Criterion, or DIC. Heuristically, these criteria try to weigh the complexity of the model against its ability to explain the data. In very general terms, if the training data indicates that a "simple" model is sufficient, we would like the model choice procedure to select a "simple" model, and correspondingly for "complex" training data and models.

Multilevel neural networks are a very flexible class of models for regression- or classification-type settings, with the ability to represent very complex model information while still being conceptually simple and general. In particular, when used in conjunction with reinforcement learning, they may have the ability to scale their complexity, in the sense that training data with increasingly complex information can result in increasingly complex model predictions. In this sense, neural networks could be described as a kind of "generic" models, with properties adaptable to a very wide range of contexts. At the same time, model parameters have proven very difficult to interpret, so model selection based on contextual knowledge, as is common for example when specifying reasonable independencies using graphical models, may be unavailable.

In this project, our main objective is to study statistical model selection for neural network models. During the last few years, a lot of heuristics have been developed to guide the building of neural network models in a wide variety of contexts. However, the focus has often been on finding something that works for a specific problem using trial-and-error rather than describing, understanding, and generalizing the model choice process. Statistical model selection theory, and in particular information criteria, has the potential to improve neural network model building.

Understanding statistical model selection theory for neural network models is also very important from another point of view: If we consider neural networks as generic types of statistical models, any understanding of how model selection works for neural networks could quite possibly have important implications for statistical model selection in general. Crucially, as neural networks seem to have the property that they can scale the complexity of their predictions based on the complexity of their training data without a change of the basic model structure of nodes and weights, good model selection theory in this context promises to capture precisely what is the most difficult issue in model selection in general: Separating the crude dimensionality of a model from its "complexity" in a more general sense. Studying implications of neural network model selection theory for model selection theory in general is the second objective of the project.

Project content

In an effort to adress the problem of overfitting, we will focus on Bayesian neural networks, so we will require a prior density on the space of network parameters. We may then use the following notation for both neural network and other types of Bayesian models: Assume we want to predict y_0 from x_0 using a training set of pairs (x_i, y_i), $i = 1, \ldots, N$. Writing $x = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_N)$, a model M for this task establishes a parameter space Θ_M, a prior density $\pi(\theta \mid M)$ for $\theta \in \Theta_M$, and a likelihood model $\pi(y_i \mid x_i, \theta, M)$ so that the required prediction is given by the density

$$\pi(y_0 \mid x_0, x, y, M) = \int_\theta \pi(y_0 \mid x_0, \theta, M)\pi(\theta \mid x, y, M)\, d\theta$$

where

$$\pi(\theta \mid x, y, M) \propto \pi(\theta \mid M) \prod_{i=0}^{N} \pi(y_i \mid x_i, \theta, M).$$

In a first and second paper, we will explore the use of model selection criteria such as DIC for the selection of M from the set of neural networks, for specific contexts (see below). In particular, we will investigate how currently used heuristics for model selection may correspond to more theoretical statistical criteria.

In a third and fourth paper, we would for the same contexts also look at models M constructed in a more context-specific tranditional way. We would then explore the utility of model selection procedures applied to both types of models.

Project participants

A PhD student. One should aim for a candidate with a strong background in both mathematical statistics and computer science.

Proposed supervisor: Professor Petter Mostad. During 2012-2017, Mostad had funding from Vetenskapsrådet for a project focusing on inference for Bayesian Networks. The more theoretical papers from this work are or will be published in Machine Learning journals. Mostad has also worked on a wide variety of applied projects where practical use of model selection has been important.

Proposed assistant supervisor: Professor Rebecka Jörnsten. Jörnsten has a number of publications which are related to Machine Learning problems and methods. She has worked on model selection for high-dimensional regression and network models, model selection for clustering and on methods for data integration. CV is available on request.

Proposed assistant supervisor: Professor Alexander Schliep. Schliep has a strong record in statistics-related computer science. He has worked on applications related to gene expression and DNA sequence analysis, and has also made a number of more theoretical contributions, in particular in relation to MCMC methods.

Links between the project propsal and AI

Finding neural network models that actually work in a given context is arguably one of the most central issues in AI today. A lot of heuristics have been developed for a large variety of contexts. However, most work has been focused on finding models that work using trial-and-error, rather than trying to develop algorithms and theory for the model selection procedure. In this situation, contributions from a statistical point of view using theory developed for model choice for context-driven models will prove valuable, and will contribute to advancing the capabilities of neural networks.

The central issue of the project is to increase understanding of model selection. Our starting point is applying known statistical theory for model selection to the issue of selecting neural network models. However, we also want to use the general modelling properties of neural networks to understand model selection for classical context-generated Bayesian models. Thus, in this second objective, we aim to use recent advances in AI to improve statistical theory. The second goal is more general and ambitious than the first, and our initial focus will be on model selection for neural network models.

The first order of business for a PhD student in this project will be to learn how successful neural network models are currently built, focusing on one or two application areas. Considering competences that are well established at Chalmers Mathematical Sciences, examples of natural choices for subject areas are:

- Clustering and data integration.
- Image analysis, and more generally spatial statistics.
- Sequence analysis (as in DNA sequence analysis).

The membership of the student in the planned Research School will be essential, as a means to acquire required knowledge. The placement of the student at Mathematical Sciences, together with the inclusion of Alexander Schliep as an advisor from Computer Science in the project, will help the student detect, use, and communicate the interdisiplinarity of the project, with tools from both mathematical statistics and computer science coming together. The work will have consequences for both the building of neural network models as part of computer science, and model selection theory within matematical statistics. The more ambitious goal of using modelling properties of neural network models to understand and improve model selection methods for classical Bayesian context-driven models will be studied towards the end of the project.