Cyprus University of Technology
Faculty of Engineering and Technology

Doctoral Dissertation

Bayesian Inference Techniques for Deep Learning

Charalampos Partaourides

Limassol, January 2018

CYPRUS UNIVERSITY OF TECHNOLOGY

FACULTY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING, COMPUTER ENGINEERING AND INFORMATICS

Doctoral Dissertation

Bayesian Inference Techniques for Deep Learning

Charalampos Partaourides

Limassol, January 2018

Approval Form

Doctoral Dissertation

Bayesian Inference Techniques for Deep Learning

Presented by

Charalampos Partaourides

Supervisor: Dr. Sotirios Chatzis, Assistant Professor

Signature ____

Member of the committee: Dr. Elpida Keravnou-Papailiou, Professor

Signature ____

Member of the committee: Dr. Theodora Varvarigou, Professor

Signature ____

Cyprus University of Technology

Limassol, January 2018

Copyrights

I would like to thank ....

my supervisor Dr. Sotirios Chatzis without whom nothing would have been possible, my friends and family that still don't know what i do, and also the people that constitute the Cyprus University of Technology; the teachers and the students, the academic and the administrative staff. All have provided me with the constructive friction needed to cherish the PhD experience.

A few words ...

Obtaining a PhD is a daunting task. It is a treacherous journey that, more often that it should, feels like it will never end. It may be bleak to reconcile that what is produced with so much effort may be obsolete in 20 years time but that is not what the PhD has essentially offered. An undergraduate degree's ultimate goal is to provide you with a coherent thinking process in order to evaluate and judge the things we all take for granted. In contrast, a PhD's goal is to provide you with the imagination tools in order to derive new things. These may be or not be the future status quo.

As researchers that is the risk we take. Nobody knows what the future will bring and in my opinion it is better not to completely know. A partially observable environment may be surrounded with uncertainty but that uncertainty is an excellent motivator for exploration.

ABSTRACT

Deep learning has achieved state of the art performance in various challenging machine learning tasks pushing the Artificial Intelligence frontier into new heights. Tasks like object recognition, speech perception, language understanding and robotics are improving year by year. This is mainly due to the recent breakthroughs in Bayesian inference, the increased volume of datasets and the increased computational power. These make it feasible to tractably train these challenging hierarchical structured models that contain millions of parameters.

Deep Learning is an umbrella term which entails numerous deep architecture models that are able to capture even the most complex dynamics of the environment. Typically, they are trained under the maximum likelihood estimation paradigm. Unfortunately, in many real world tasks the high dimensionality of the observations results in even the largest datasets to being sparse. As such, there is an immense need for the training algorithm to compensate the uncertainty introduced by the data sparsity, overcome the model's overfitting tendencies and in result generalize well.

The statistical method of Bayesian inference provides a mathematically coherent way of dealing with data sparsity and overfitting. It essentially uses the Bayes theorem to accumulate evidence-based knowledge. This is achieved by postulating probability distributions over the parameters instead of trying to derive point estimates of them. Under the Bayesian view, we impose a prior distribution that encapsulates our initial belief about the model's dynamics and we correct that belief as we are presented with more data; this consists in inferring the posterior distribution. It is conspicuous that the choice of the distribution heavily controls the expressiveness of the model.

In this thesis, we present innovative approaches to train deep networks by considering sparsity, skewness and heavy tails on the form of the parameters distribution. Specifically, among our contributions, we impose a sparsity inducing distribution over the network synaptic weights to improve generalization. On a different vein, we consider the imposition of a skew normal distribution over the latent variables to increase the deep networks capacity. In parallel, we examine the efficacy of inferring the feature functions by devising a novel random sampling rational combined by an optimizable sample weighting scheme. The models derived by the aforementioned approaches are trained by means of approximate Bayesian inference scheme to allow for scalability in large datasets. We exhibit the advantages of these methods over existing approaches by conducting an extensive experimental evaluation using benchmark

datasets.

Keywords: Deep Learning, Machine Learning, Bayesian Inference, Variational Bayes, Regularization