View Reviews

Paper ID
6866

Paper Title
Learning Exponential Families

Reviewer #1

Questions

1. Please provide an "overall score" for this submission.
5: Marginally below the acceptance threshold. I tend to vote for rejecting this submission, but accepting it would not be that bad.

2. Please provide a "confidence score" for your assessment of this submission.
2: You are willing to defend your assessment, but it is quite likely that you did not understand central parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

3. Please provide detailed comments that explain your "overall score" and "confidence score" for this submission. You should summarize the main ideas of the submission and relate these ideas to previous work at NIPS and in other archival conferences and journals. You should then summarize the strengths and weaknesses of the submission, focusing on each of the following four criteria: quality, clarity, originality, and significance.
The authors propose a mechanism to approximate an intractable exponential family model as opposed to a finding specific representative of some intractable family. The motivation is to learn expressive models whose posteriors are easy to calculate, hence making them useful for inference. I found the paper to be somewhat difficult to read, owing partly to my lack of familiarity with the related work but also as a result of the wordiness of some of the descriptions. It is also a bit too colloquial in parts. As an example of wordiness without precision: " well-known are the overfitting/generalization associated with a finite dataset compared with access to a distribution..." It's better just to list some and be precise rather than to hope that these things really are well known to the reader. Finally, the experiments are interesting showcase a wide range of behaviors that these kinds of methods can exhibit.

Specific questions:

line 108, " Though trivial to sample from qθ(z) for any choice of family G"... Why is this true? I don't think that it is easy to sample from arbitrary exponential family models. Unless I'm missing something here?

What is q_0 at the end of line 108?

What is \Phi in line 122? Is F just some arbitrary parameterized family of maps from the natural parameters H into a \Theta?

I don't understand the notation in line 130 as it hasn't been defined. I assume that you mean q_{f_{\phi}(\eta)}(z)? This notation changes again on line 133 :(

Line 140 makes it sound like the only meaningful objective here requires sampling, that seems strange to me. For example, you couldn't you also minimize over \phi and then maximize over \eta?

I don't like the way equation (2) is presented. I think you should explicitly state that \eta_1,...,\eta_K are sampled from p(.), etc.

Typos:

"classic" -> "classical" almost everywhere this is used

line 144, "computing calculating" -> "calculating"

line 161, need a \in

line 201, "vs" -> "versus" (appears later as well)

4. How confident are you that this submission could be reproduced by others, assuming equal access to data and resources?
1: Not confident

Reviewer #2

Questions

1. Please provide an "overall score" for this submission.
5: Marginally below the acceptance threshold. I tend to vote for rejecting this submission, but accepting it would not be that bad.

2. Please provide a "confidence score" for your assessment of this submission.
4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

3. Please provide detailed comments that explain your "overall score" and "confidence score" for this submission. You should summarize the main ideas of the submission and relate these ideas to previous work at NIPS and in other archival conferences and journals. You should then summarize the strengths and weaknesses of the submission, focusing on each of the following four criteria: quality, clarity, originality, and significance.
Summary
==========

The paper focus on learning an exponential family to model the posterior distribution. While most of the methods look at the problem of learning a single posterior distribution, they propose to learn a parameterized family of distributions. In this way, they assume that the posterior is given by an intractable exponential family (P) and they aim at learning a family of distributions (Q) that approximate P by minimizing the KL divergence. Therefore, the training procedure consists in learning a region in the space of the parameters indexing Q. The paper is well-written and the method well-motivated. The explanation is clear but the experimental section needs more work. In particular, the experiments fail to explain some basic aspects of the proposal (underperformance of EFN vs EFN1, differences between EFN1 and NF1, …) and they do not showcase some of the selling points of the proposal, e.g. the application to meta learning. Finally, some experiments with real datasets would improve the quality of the paper.

Details
==========

The authors focus on learning a family of distributions (Q) that approximate the intractable posterior exponential family (P). To do so, they use a “density network” (a bijective mapping between a random variable and the output of the posterior distribution). Unlike previous papers that aim at learning a single member (that is close to the true posterior) inside the family implicitly defined by the parameters of the “density network”, they use a second network called the “parameter network” to model the parameters of the “density network”. Therefore, the output is no longer a single distribution, but a family defined by a subspace of the parameter space of the “density network” defined by the learnt parameters of the “parameters network”. In this sense, the proposed approach is similar to previous work in meta-learning.

In the experimental section they firstly apply the algorithm to a set of tractable distributions. They compare their approach (EFN) to EFN1 and NF1: two baselines that learn a point-wise distribution rather than a family of distributions. The experiments a well-motivated and clear, however, they fail to explain some basic aspect of the algorithm. In particular, their method (EFN) seems to perform worse that EF1, which is equal to EFN but with a fix value for the natural parameters of the posterior, and therefore, it outputs a single posterior distribution. This does not support their initial hypothesis: that learning family of distributions is better than learning a single distribution to represent the posterior. The authors claim that the performance between EFN and EFN1 is quite similar and that EFN could be used to initialize EFN1, however, they do not explore further this proposal. Also they make connections to meta-learning and highlight some potential advantages of the method in this area but they do not propose experiments in this line.

In section 3.2 they apply the algorithm to a hierarchical dirichlet model. Here they try to analyze the difference between EFN1 and NF1 (same as EFN1 but removing the parameters network). The analysis is interesting but they do not reach a conclusion that explain the difference in performance. Further analysis needs to be done. Finally, they show some interesting results in section 3.3 showing how the proposed method outperforms VI.

All the experiments are done with synthetic datasets. It would be desirable to add experiments with real datasets that show the benefits of the proposed approach clearly.

4. How confident are you that this submission could be reproduced by others, assuming equal access to data and resources?
3: Very confident

Reviewer #3

Questions

1. Please provide an "overall score" for this submission.
4: An okay submission, but not good enough; a reject. I vote for rejecting this submission, although I would not be upset if it were accepted.

2. Please provide a "confidence score" for your assessment of this submission.
4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

3. Please provide detailed comments that explain your "overall score" and "confidence score" for this submission. You should summarize the main ideas of the submission and relate these ideas to previous work at NIPS and in other archival conferences and journals. You should then summarize the strengths and weaknesses of the submission, focusing on each of the following four criteria: quality, clarity, originality, and significance.

The paper proposes a variational inference method for exponential family (EF) models by approximating the variational distribution as a two-neural network architecture. One neural network is used to tie the (natural) parameters of EF models to be learned with the parameter of the recognition network.
The flow of the paper is hard to follow as its motivation is not clearly stated. How does the current work differ from the existing fundamental works for learning EF models such as VB, SVI, VAE?
In Eq. (2), the estimator uses the inverse function of g_\theta(.), how to compute this function value when you have only g_\theta(.)? Is the NN used to approximate g^{-1}_\theta(.) not g_\theta()?
The experiments are not comprehensive and convinced enough to prove the efficiency of the proposed method. The experiments demonstrated some number from alternatives of the proposed method. No baseline methods such as VB, SVI, or VAE are used to compare with the proposed inference mechanism.

4. How confident are you that this submission could be reproduced by others, assuming equal access to data and resources?

2: Somewhat confident