2500 boul. de l'Université
Sherbrooke, QC
J1K 2R1

November 2nd 2013

Thesis evaluation report

To whom this may concern,

I hereby present my evaluation of Kyunghyun Cho's thesis, title "Deep Neural Networks : Basic Principles and Recent Advances".

The thesis makes several contributions to the popular field of deep learning. These contributions have been published across 10 different publications, which makes for an impressive tally. Two of them have been published at the ICML conference, one of the top conferences in the field. It is thus undenieable that the PhD candidate has made a sufficient amount of meaningful contributions to obtain his degree.

In summary, the candidate has proposed several ways of improving the training of different variants of Boltzmann machines (mainly restricted Boltzmann machines and deep Boltzmann machines). He has suggested an improved descent direction for gradient optimizaiton of RBMs (the *enhanced gradient*), an adaptive learning approach, a parallel-temporing-based stochastic optimization algorithm and a new pre-training approach for deep Boltzmann machines. In addition, he proposed a novel sparsification approach for denoising autoencoders and presented the applications of deep learning methods to image denoising and speech recognition.

In addition to the many papers that describe these contributions, the candidate has also put together an exhaustive literature review of deep learning, in which he inserted his own work. A lot of work has been put in this review and I've really enjoyed reading it. I'm seriously considering using some of it as reading material for my own neural network course.

Given that the literature review is the only material in this thesis that hasn't yet gone through peer review, my comments about the thesis focus on it. I have a few suggestions to make to the candidate, to improve the review.

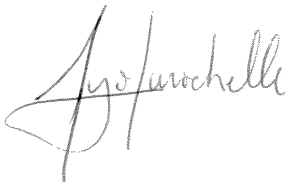
These will require mainly minor changes :

– I'm not a big fan of section 3.2.2, mainly because I feel that the link between variational inference and the encoder of an autoencoder is not very strong. Specifically, an encoder isn't trying to reduce the KL divergence between its output (interpreted as a distribution) and the true posterior of a latent variable model. It is instead trained to reduce reconstruction error, which isn't the same thing. I think this additional caveat should at least be mentioned in the section, and some emphasis should be put on the fact that matching the parametric form of some variational posterior and performing variational inference are two different things (where variational inference is also associated with a specific training objective for the variational posterior).
– The description of the DBM pretraining procedure in section 5.3.3 is not clair enough, specifically the justification for why it improves the variational bound. I should say that I myself have never fully understood the argument, and I don't think the Salakhutdinov and Hinton paper doesn't do a great job at proving the variational bound improvement. Thus there is an opportunity here for providing a clearer explanation. Specifically, the questions that remain, after reading this section are :
    1. Why should we train the first RBM with two copies of the input, i.e. why should we treat $\mathbf{h}^{[2]}$ as if it were observed and exaclty equal to $\mathbf{x}$ in the training set ?
    2. Is the bound tihgt at any point, e.g. when $\theta_1 = \theta_2$? I don't think it is but, if it should be, an explanation should be given.
    3. What does it mean that the $p(\mathbf{h}^{[1]}|\theta_1, \theta_2)$ has a smaller KL-divergence **than** the aggregate posterior ? Do you mean **with** the aggregate posterior ?
– Overall, I've identified a lot of little English mistakes throughout the thesis. I'm attaching my annotated copy of the thesis, so the candidate can make the appropriate corrections. Moreover, I make other, more minor comments about the thesis in the annotated copy, which the candidate can consider to further improve the thesis.

Overall, I think this is an excellent thesis, that belongs to the 10% best thesis I've had a chance to examine. I thus recommend the thesis be accepted after

the minor corrections mentioned above have been made.

Respectfully yours,



Hugo Larochelle