TUNING DBN HYPER-PARAMETERS USING A GENETIC ALGORITHM

AUTHORS TBD

1. Introduction

Deep belief networks (DBNs) are neural networks which are first pre-trained in a greedy layer wise manner using Boltzmann learning and then fine tuned using back-propagation. DBNs implement state of the art image and speech recognition. While the results can be impressive,

The significant increase in performance of these DBNs over traditional neural networks comes at a cost of an increase of hyper-parameters which must be tuned to ensure even good performance. The process of tuning these hyper-parameters can be tedious, time consuming and expensive. Even worse, whether a set hyper-parameters works well also may depend on what hardware is used to train the classifier.

2. Laundry list of hyper-parameters

Random Seed
Classifier learn rate
Classifier l2-penalty
Classifier l1-penalty
Number of layers (n. layers)
Number of hidden units (n. hidden-units)
(Mini) Batch size
Weight initialization method
CD epochs
CD learn rate
Convolution networks: (additional constraint on the weights)?
Cost Function
Layer by layer fitness function evaluation
Activation function
Pre processing strategy: PCA, Fourier Transform, et al.
Defining the hyper-parameters:

As identified Supra. We will define each hyper parameter to clarify the process in which its assigned.

Random Seed a selection of random numbers or vectors that is used to generate pseudo-random numbers.

Classifier Learn Rate how quickly a gradient moves when updating any given weight.

Classifier L2-penalty a term that is added to the cost function to ensure regularization.

Classifier L1-penalty a second term that is added to the cost function to ensure regularization.

Number of layers Roughly corresponds to the level of hierarchical abstraction for the overall network.

Number of hidden units Roughly corresponds to the amount of features that may be processed in each layer.

Date: July 12, 2014.

(Mini) Batch size The number of samples selected from the trading set to evaluate the gradient decent at any given time.

Weight initialization method procedure (WIMP) Strategy for introducing the initial weights.

CD Epochs The maximum number of iteration during Boltzmann optimization.

CD Learn Rate Learning right for the Boltzmann optimization.

Convolution networks - Whether to use or not to use a convolutional network (using one introduces a whole new set of hyper-parameters).

Cost Function Which type of cost function to use, such as: logistic, linear, etc.

Layer by layer fitness function evaluation Whether to use it .

Activation Function Such as: sigmoid, arc tan, etc.

Pre Processing Strategy whether to use raw data or to pre process it using: PCA, Fourier Transform, et al.

3. Genetic algorithm and multi-objective optimization

4. Methods

5. Results

6. Future work

7. Reading

References

1. Hussein A Abbass, *Speeding up backpropagation using multiobjective evolutionary algorithms*, Neural Computation **15** (2003), no. 11, 2705–2726.
2. Sanghamitra Bandyopadhyay and Sankar Kumar Pal, *Classification and learning using genetic algorithms: applications in bioinformatics and web intelligence*, Springer, 2007.
3. Rémi Bardenet, Mátyás Brendel, Balázs Kégl, Michele Sebag, et al., *Collaborative hyperparameter tuning*, JMLR Workshop and Conference Proceedings, vol. 28, 2013, pp. 199–207.
4. Yoshua Bengio, *Gradient-based optimization of hyperparameters*, Neural computation **12** (2000), no. 8, 1889–1900.
5. ———, *Practical recommendations for gradient-based training of deep architectures*, Neural Networks: Tricks of the Trade, Springer, 2012, pp. 437–478.
6. James Bergstra, Rémi Bardenet, Yoshua Bengio, Balázs Kégl, et al., *Algorithms for hyper-parameter optimization.*, NIPS, vol. 24, 2011, pp. 2546–2554.
7. James Bergstra and Yoshua Bengio, *Random search for hyper-parameter optimization*, The Journal of Machine Learning Research **13** (2012), 281–305.
8. James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio, *Theano: a CPU and GPU math expression compiler*, Proceedings of the Python for Scientific Computing Conference (SciPy), June 2010, Oral Presentation.
9. Ingrid Daubechies et al., *Ten lectures on wavelets*, vol. 61, SIAM, 1992.
10. Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan, *A fast and elitist multiobjective genetic algorithm: Nsga-ii*, Evolutionary Computation, IEEE Transactions on **6** (2002), no. 2, 182–197.
11. Deeplearning.net, *Deeplearning*, `http://deeplearning.net/software/theano/tutorial/index.html`, Accessed: 2014-02-10.
12. Richard O Duda, Peter E Hart, and David G Stork, *Pattern classification*, John Wiley & Sons, 2012.
13. David Edward Goldberg et al., *Genetic algorithms in search, optimization, and machine learning*, vol. 412, Addison-wesley Reading Menlo Park, 1989.
14. Benjamin Graham, *Sparse arrays of signatures for online character recognition*, arXiv preprint arXiv:1308.0371 (2013).
15. John J Grefenstette, *Optimization of control parameters for genetic algorithms*, Systems, Man and Cybernetics, IEEE Transactions on **16** (1986), no. 1, 122–128.
16. Geoffrey Hinton, *Neural nets*, `http://class.coursera.org/neuralnets-2012-001/`, Accessed: 2014-02-10.
17. ———, *A practical guide to training restricted boltzmann machines*, Momentum **9** (2010), no. 1, 926.

18. Yaochu Jin, Tatsuya Okabe, and Bernhard Sendhoff, *Neural network regularization and ensembling using multi-objective evolutionary algorithms*, Evolutionary Computation, 2004. CEC2004. Congress on, vol. 1, IEEE, 2004, pp. 1–8.
19. C Richard Johnson, Ella Hendriks, Igor J Berezhnoy, Eugene Brevdo, Shannon M Hughes, Ingrid Daubechies, Jia Li, Eric Postma, and James Z Wang, *Image processing for artist identification*, Signal Processing Magazine, IEEE **25** (2008), no. 4, 37–48.
20. Lamos-Sweeney Joshua, *Deep learning using genetic algorithms*, Master's thesis, 2012.
21. Kaggle.com, *Cifar-10 - object recognition in images*, `http://www.kaggle.com/c/cifar-10/`, Accessed: 2014-02-10.
22. Alex Krizhevsky and Geoffrey Hinton, *Learning multiple layers of features from tiny images*, Computer Science Department, University of Toronto, Tech. Rep (2009).
23. Alex Krizhevsky, Geoffrey E Hinton, et al., *Factored 3-way restricted boltzmann machines for modeling natural images*, International Conference on Artificial Intelligence and Statistics, 2010, pp. 621–628.
24. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, *Imagenet classification with deep convolutional neural networks.*, NIPS, vol. 1, 2012, p. 4.
25. Frank Hung-Fat Leung, Hak-Keung Lam, Sai-Ho Ling, and Peter Kwong-Shun Tam, *Tuning of the structure and parameters of a neural network using an improved genetic algorithm*, Neural Networks, IEEE Transactions on **14** (2003), no. 1, 79–88.
26. David JC MacKay, *Information theory, inference and learning algorithms*, Cambridge university press, 2003.
27. Andrew Ng, *Machine learning*, `http://www.cousera.org/course/ml/`, Accessed: 2014-01-05.
28. Mohammad Norouzi, *Convolutional restricted boltzmann machines for feature learning*, Master's thesis, School of Computing Science-Simon Fraser University, 2009.
29. SN Sivanandam and SN Deepa, *Genetic algorithm optimization problems*, Springer, 2008.
30. Jasper Snoek, Hugo Larochelle, and Ryan P Adams, *Practical bayesian optimization of machine learning algorithms.*, NIPS, 2012, pp. 2960–2968.
31. Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus, *Intriguing properties of neural networks*, arXiv preprint arXiv:1312.6199 (2013).
32. Vikrant Singh Tomar and Richard C Rose, *Efficient manifold learning for speech recognition using locality sensitive hashing*, Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp. 6995–6999.
33. Aimin Zhou, Bo-Yang Qu, Hui Li, Shi-Zheng Zhao, Ponnuthurai Nagaratnam Suganthan, and Qingfu Zhang, *Multiobjective evolutionary algorithms: A survey of the state of the art*, Swarm and Evolutionary Computation **1** (2011), no. 1, 32–49.