SPECIAL TOPICS IN ADVANCED MACHINE LEARNING
SPRING 2017
INSTRUCTOR: ANNA CHOROMANSKA

LIST OF PROJECTS

Topic 1 and 2 (Deep learning):

Modern state-of-the-art convolutional neural networks have hundreds of millions of parameters that need to be stored in memory. Mobile devices and memory constrained devices often have insufficient capacity to store such networks. Familiarize yourself with the related literature on network compression (weight quantization, structured parametrization etc.). Read only top-cited papers, you can search for them in Google Scholar. Review the literature you read. Try to propose a new compression method. Explore empirically different ways of compressing the network. Optionally, you can also try to work on the theoretical guarantees of a proposed method. BONUS: Put this network on an actual mobile device and demonstrate its performance.

Topic 3 and 4 (Deep learning):

Adversarial networks are extremely valuable deep learning models used nowadays for a number of interesting problems. Familiarize yourself with the literature on adversarial networks. Start from Goodfellow et al., Generative adversarial nets, NIPS, 2014. Analyze different learning frameworks and applications when these models are used in the literature. Review the literature you read. Read only top-cited papers, you can search for them in Google Scholar. Propose a new problem and architecture to which adversarial setting can be applied. Justify your choice with a discussion and empirical evaluation.

Topic 5 (Deep learning)

Consider a 10-class classification problem based on MNIST dataset (a collection of images of digits). Consider two different ways of regularizing the network. The first one is dropout. To learn about it read the paper of Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, Sutskever, Ilya, and Salakhutdinov, Ruslan. Dropout: A simple way to prevent neural networks from overfitting. JMLR, 15:1929–1958, 2014. Review this paper. The other regularization is the following: when you train the network instead of using binary vectors as labels, use vectors that add uncertainty to your prediction. For example, image of digit 3 has a label assigned to it of the form (0,0,0,1,0,0,0,0,0,0,0) - it is a vector of legth 10 (since a network has 10 outputs) when the i-th entry in this vector correspond to digit i-1 (i = 1,2,3,...,10). Instead to train the network, use the label vector of the form (a/9,a/9,a/9,1-a,a/9,a/9,a/9,a/9,a/9,a/9), where a belongs to interval [0,0.5). For example a=0.05. Try also randomized version of this scheme. Describe the effect of this new regularization for different settings of a and empirically compare both regularizations on MNIST and perhaps other datasets. Optionally, you can also try to work on the theoretical description of the second regularization technique.

Topic 6 (Optimization):

Familiarize yourself with the majorization method as proposed in T. Jebara, A. Choromanska, Majorization for CRFs and Latent Likelihoods, in the Neural Information Processing Systems Conference (NIPS), 2012 (read first 3 pages). Review this paper. The method is proposed in the batch setting. Explore adapting the method to the stochastic setting. You may want to use the work A. Choromanska, T. Jebara, Stochastic bound majorization, CoRR, abs/1309.5605, 2013, 2013 as a guideline. Either propose a different stochastic version of the bound majorization method than the one shown in this paper or rely on the one described there. Provide empirical evaluation of the proposed method (compare with other optimization tools) or work on the theoretical description of the method, e.g. convergence guarantees, or both.

Topic 7 (Optimization):

Familiarize yourself with the majorization method as proposed in T. Jebara, A. Choromanska, Majorization for CRFs and Latent Likelihoods, in the Neural Information Processing Systems Conference (NIPS), 2012 (read first 3 pages). Review this paper. Explore the method in the sparse learning setting for sparse logistic regression (logistic regression with L1-norm penalty). Provide empirical evaluation of the proposed method (compare with other optimization tools for sparse optmization). Optionally, you can also try to work on the theoretical description of the method.

Topic 8 and 9 (Optimization):

Familiarize yourself with the majorization method as proposed in T. Jebara, A. Choromanska, Majorization for CRFs and Latent Likelihoods, in the Neural Information Processing Systems Conference (NIPS), 2012 (read first 3 pages). Study if the method can achieve the super-linear convergence rate. Perhaps Theorem 3.1. in

https://www.math.washington.edu/~burke/crs/408/notes/Math408_Spring2014/directions-unco.pdf

is a right tool to use.

Topic 10 and 11 (Multi-class and multi-label classification):

Familiarize yourself with the literature on multi-label classification and also the work of Anna Choromanska and John Langford, Logarithmic time online multiclass prediction, CoRR, abs/1406.1822, 2014. Review the literature you read. Read only top-cited papers, you can search for them in Google Scholar. Modify the algorithm in Anna Choromanska and John Langford, Logarithmic time online multiclass prediction, CoRR, abs/1406.1822, 2014 to perform multi-label classification or propose yet a completely new multi-label classification algorithm. Perform empirical evaluation of your method. Optionally, you can also try to work on the theoretical description of the proposed method.

Topic 12 (Multi-class classification):

Familiarize yourself with the analysis of multi-class classification as given in A. Choromanska, K. Choromanski, M. Bojarski, On the boosting ability of top-down decision tree learning algorithm for multiclass classification, CoRR, abs/1605.05223, 2016 as well as with the paper M. Kearns and Y. Mansour. On the boosting ability of top-down decision tree learning algorithms. Journal of Computer and Systems Sciences, 58(1):109–128 (also In STOC, 1996), 1999. Review these papers. Consider the multi-class classification objective J(h) as given in Equation 1 in the first paper. This paper shows how fast entropy-based criteria are reduced when we optimize objective J(h) in each node of the classification tree. Pick one entropy-based criterion from this paper. Can you propose a theoretical framework to directly compare J(h) with this entropy-based criteria?

Topic 13 (Multi-class classification):

Familiarize yourself with the analysis of multi-class classification as given in A. Choromanska, K. Choromanski, M. Bojarski, On the boosting ability of top-down decision tree learning algorithm for multiclass classification, CoRR, abs/1605.05223, 2016. Review this paper. Consider the multi-class classification objective J(h) as given in Equation 1 in this paper. This paper shows how fast entropy-based criteria are reduced when we optimize objective J(h) in each node of the classification tree. Can you modify the theoretical framework in this paper or propose a new framework to show how fast the multi-class classification error is reduced when we optimize the binary classification error in each node of the classification tree?

Topic 14 (Bounding strategies)

Familiarize yourself with the paper Piecewise Bounds for Estimating Bernoulli-Logistic Latent Gaussian Models, (ICML 2011) B. Marlin, M. E. Khan, K. Murphy and also T. Jebara, A. Choromanska, Majorization for CRFs and Latent Likelihoods, in the Neural Information Processing Systems Conference (NIPS), 2012 (read first 3 pages of the latter). Review both papers. Note that the first one explores different ways of bounding the logistic-log-partition function, and the authors of this paper eventually choose to use piecewise bounds. Modify their approach such that instead of using these bounds, you will use the bounding scheme from T. Jebara, A. Choromanska, Majorization for CRFs and Latent Likelihoods, in the Neural Information Processing Systems Conference (NIPS), 2012. Describe the new approach. Compare empirically both approaches (the code for the first approach can be found in http://icapeople.epfl.ch/mekhan/publications.html). Optionally, you can also try to work on the theoretical description of the new method.

Topic 15 (Stability of learning algorithms)

Read the paper of Olivier Bousquet and Andre Elisseeff, Stability and generalization, Journal of Machine Learning Research, 2 (Mar):499–526, 2002. It introduces the notion of the stability of learning algorithms. Read also the paper on clustering stability of U. von Luxburg, Clustering stability: An overview, Foundations and Trends in Machine Learning, 2(3):235–274, 2010. Review both papers. Design some experiments illustrating the notion of stability in settings of your choice.

Topic 16 and 17 (Experts advice)

Read the paper of M. Herbster and M. K. Warmuth. Tracking the best expert. Machine Learning, 32:151–178, 1998 dedicated to performing prediction based on experts advice in a supervised learning setting (you may find chapters 2 and 3 in http://faculty.cs.gwu.edu/~cmontel/AITR-2003-011.pdf very useful too). Review this paper and the papers that use the experts advice in different learning settings (e.g.

A. Choromanska, C. Monteleoni, Online clustering with experts, AISTATS, 2012 adapt expert advice approach to the unsupervised learning setting - clustering). Can you propose a learning setting where experts advice was not yet used but can be potentially be useful? Work on adapting the experts advice to this setting. You can explore combining deep learning setting with the experts framework. Optionally, provide empirical evaluation of the proposed method or work on the theoretical description of the method, or both.

Topic 18 (Clustering):

Explore the connection between kernel k-means and spectral clustering as given in Dhillon, Inderjit S., Guan, Yuqiang, Kulis, Brian, 2004. Kernel k-means: Spectral clustering and normalized cuts. In: Proc. 10th KDD, pp. 551–556. Read also the papers on online k-means of E. Liberty, R. Sriharsha, and M. Sviridenko. An algorithm for online k-means clustering. CoRR, abs/1412.5721, 2014 and A. Choromanska, C. Monteleoni, Online clustering with experts, AISTATS, 2012. Review these papers. Can you propose an online spectral clustering algorithm that results from the online k-means algorithm by using the connection between kernel k-means and spectral clustering? Compare your algorithm with other incremental spectral clustering techniques with a discussion and empirical evaluation. Optionally, you can also try to work on the theoretical description of the proposed method.

Topic 19 (Clustering):

Explore the connection between kernel PCA and spectral clustering: find the literature related to this problem, e.g. see Bengio, Y., Vincent, P., & Paiement, J. F. (2003). Learning eigenfunctions of similarity: linking spectral clustering and kernel PCA, Technical Report, University de Montreal (see also technical report Spectral Clustering and Kernel principal component analysis are pursuing good projection, V. Chandrakant Raykar), etc. Furthermore, find the literature on Online Principal Component Analysis. Review the papers you read. Read only top-cited papers, you can search for them in Google Scholar. Can you propose an online spectral clustering algorithm that results from the online PCA algorithm by using the connection between kernel PCA and spectra clustering? Compare your algorithm with other incremental spectral clustering techniques with a discussion and empirical evaluation. Optionally, you can also try to work on the theoretical description of the proposed method.

List of popular machine learning datasets:

- MNIST
- CIFAR-10 and CIFAR-100
- SVHN
- ImageNet
- UCI Machine Learning Repository: Data Sets
- LIBSVM Data