Homework set - Neural networks - Part 2

September 16, 2018

Guidelines

- For the use of this assignment we use the MNIST dataset[1].
- The solution for this homework is to be posted as a .pdf file.
- You may choose the programming language you prefer for the implementations (excluding packages that enable auto-differentiation, e.g tensorflow, keras etc.).
- All plots must have named axis, grids and title. If more than one plot is on the same figure, provide legend.
- Use your best model from previous homework (Neural networks - Part 1) as a baseline model.

1 Introduction

In this homework, we survey regularization methods of neural networks. The goal of the regularization is to restrict the model during training in order to achieve lower generalization error (less over-fitting).

2 Self-Reading

Read and solve the exercises in chapter 3 at Michael Neilsen e-book at the following link.

3 Regularize your best model

In this section you should implement at least **three** from the following regularization methods over your **best** model from the previous homework. It is **very** recommended to add more regularization methods (which are not on the list) that you wish to try.

1. L1, L2 regularization[1].
2. Dropout[2].
3. Batch normalization[3].
4. Confidence-penalty/label-smoothing[4]
5. Activation regularization[5]

[1] The MNIST dataset could be downloaded from this link

Submission

The work in this assignment should be summarized into a pdf document, where you present the exercises from Michael Nielsen book and the results from section 3. Please provide results over your regularized models and compare it to your base model from previous homework. The results should be presented as follows:

1. Present the code of the regularization methods you chose.
2. Document your best model as requested in part 1 of this assignment.
3. Present an ablation analysis of the regularization methods of your best model. That is, compare your best model accuracy and negative-log-loss with an ablated version of your model. E.g, if your best model was regularized with L1 regularization, dropout and batch normalization, you should compare it to a model that uses L1 regularization and dropout, a model with L1 regularization and batch normalization and a model with dropout and batch normalization.

Moreover, think of visualizations that you could use to give sense of your model's behavior. You can use ideas from this link.

GOOD LUCK!!!

References

[1] A. Krogh and J. A. Hertz, "A simple weight decay can improve generalization," in *Advances in neural information processing systems*, 1992, pp. 950–957.

[2] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[3] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167

[4] G. Pereyra, G. Tucker, J. Chorowski, Ł. Kaiser, and G. Hinton, "Regularizing neural networks by penalizing confident output distributions," *arXiv preprint arXiv:1701.06548*, 2017.

[5] S. Merity, N. Shirish Keskar, and R. Socher, "Regularizing and Optimizing LSTM Language Models," *ArXiv e-prints*, Aug. 2017.