SPARSIFICATION OF NEURAL NETWORKS WITH VARIATIONAL DROPOUT

A PREPRINT

Ilia Luchnikov
luchnikovilya@gmail.com

Marsel Faizullin
marsel.faizullin@skoltech.ru

Max Blumental
bvmaks@gmail.com

October 26, 2018

ABSTRACT

In this project we apply variational dropout to a fully connected neural network in order to sparsify it. The idea for the project is taken from "Variational Dropout Sparsifies Deep Neural Networks" article by Dmitry Molchanov and Arsenii Ashukha.

***K*eywords** Variational Dropout, Sparsity

1 Introduction

We considered classification on MNIST dataset. In our experiments we had 2- and 3-layers fully connected networks with Bernoulli, Gaussian and Variational Gaussian dropouts. We implemented variational dropout in the way described in the above mentioned article in order to achieve sparsity.

1.1 Architectures

At the table below are represented 3 architectures which we implemented.

	Architecture	Activations	Initial weights for θ	Initial weigths for $log\ \sigma^2$	Learning rate and optimizer type	Batch size
Bernoullian and Gaussian dropouts	28*28 x 256 256 x 10	ReLu, Softmax	Uniform $[-\frac{1}{\sqrt{n_{out}}}, \frac{1}{\sqrt{n_{out}}}]$	–	0.01	100
Variational dropout (1-st NN)	28*28 x 256 256 x 10	ReLu, Softmax	Uniform $[-\frac{1}{\sqrt{n_{out}}}, \frac{1}{\sqrt{n_{out}}}]$	-25	0.01	100
Variational dropout (2-nd NN)	28*28 x 100 100 x 40 40 x 10	ReLu, Relu, Softmax	Normal $[0, \frac{1}{\sqrt{n_{out}}}]$	-11.5	0.0003	100

Figure 1: Summary of architectures

1.2 Results

Here is the list of takeaways from the project:

- Variational dropout actually causes sparsity (see animation on Github).
- There is a tradeoff between sparsity and prediction accuracy. We can manage it varying α.
- It is hard to make the training process stable. Whatever initialization and numerical tricks you do, after some sufficiently large number of iterations you will get `NaN`.

1.3 Project development

- Consider other neural architectures.
- Check how variational dropout prevents networks from overfitting.
- Implement structured pruning.

2 Contribution of participants

1. Ilia Luchnikov: 3-layer NN with variational dropout, visualization of sparsity.
2. Marsel Faizullin: Implementation and comparison of Bernoullian and Gaussian Dropouts
3. Maxim Blumental: Implemented 2-layer FC network with variational dropout using authors' tricks. Visualized sparsity in my notebook.