
FAST DROPOUT TRAINING

PAPER REVIEW

Soumik Dhua

Department of Computer Science
Birla Institute of Technology and Science, Pilani
Goa - 403726
f20180155@goa.bits-pilani.ac.in

July 29, 2019

1 Introduction

Dropout is a technique widely used for preventing overfitting while training deep neural networks. The basic idea behind dropout is very simple. During the training process, in each iteration of optimization, the input features and hidden units are randomly dropped. This randomized regularization method for training deep neural networks has led to increase in predictive accuracy. This dropout technique has played an important role in models such as AlexNet for ImageNet classification. This technique can also be seen as averaging over many neural networks with shared weights.

2 Objective

Applying dropout to a deep neural network is computationally expensive and also time-consuming. This paper derives an analytic approximation to dropout. The authors give an insight into how to use all the data efficiently and achieve the benefit of dropout training. This approximation gives roughly ten times speedup in training time under certain conditions, while also giving more stability.

3 Model and Assumptions

The key understanding in this fast dropout training paper is that if the input to each node in a neural network is a weighted sum of its inputs – and if some of those inputs are being randomly dropped, then the total input becomes the weighted sum of Bernoulli random variables. If a neuron has many inputs having a comparable variance then by using the Central Limit Theorem, we can very well approximate the weighted sum of Bernoulli random variables by a Gaussian. The mean and variance of these Gaussians are integrated over all the combinations of dropouts in a one-layer network. The authors approximate the output of each neuron with a Gaussian and derive deterministic update rules for the multi-layer network.

4 Result and Conclusion

The fast dropout approximation given in this paper proves to be stable and also provides an order of magnitude speedup in training time under certain conditions. After reading this paper, we can think about dropout in two different ways– as an approximate variational method, and as a scale-invariant regularizer.

References

- [1] Sida I. Wang and Christopher D. Manning *Fast dropout training*. <https://nlp.stanford.edu/pubs/sidaw13fast.pdf>