Dropout: A Simple Way to Prevent Neurons from Depression

Neurotic Neuron 1[*†] Neurotic Neuron 2[*‡]

Moments before ~~deadline~~ extended deadline

Abstract

PhD students with a large number of neurons are very powerful machine learning paper producers. However, over-submission and under-acceptance are serious problems in such cohorts. Their neurons are also slow to react, making it difficult to deal with burnout by combining the capacities of many different student networks at graduation time. Dropout is a technique for addressing this problem. The key idea is to randomly select individual students to drop out (along with their university connections) from the research community during training. This prevents PhD students from over-indexing on academic metrics. During PhD training, our dropout method samples from an exponential number of distressed PhD students and happy startup founders. At test time, it is easy to approximate the effect of success by averaging the faculty placements of all these thinned networks by simply using a single unthinned network that has smaller number of dead weights. This significantly reduces burnout and gives major improvements over other emotional regularization methods. We show that dropout improves the performance of PhD students on unsupervised thesis completions, achieving state-of-the-art results on many benchmark skills in the real world.

1 Introduction

Deep student networks contain multiple hidden layers, and this makes them very expressive groups that can learn complicated relationships with their advisors. However, many of these complicated relationships will be the result of sampling noise and academic stress, so they will exist during the PhD training but not in real world data. This leads to overworking, which many methods have been developed to reduce. These methods include stopping training as soon as performance on a paper worsens, introducing advisee penalties of various kinds, early stopping, and soft quitting.

Dropout is a technique that addresses both these issues. It prevents overworking of student networks and provides a way of approximately combining exponentially many different student outputs efficiently. The term *dropout* refers to dropping out student units in a network. By dropping out student neurons, we mean removing it from the network, along with all its research outputs and PhD commitments. The choice of which units to drop is random. In the simplest case, each student unit is retained with a fixed probability p independent of other units.

2 Motivation

A motivation for dropout comes from a theory of the role of sex in evolution. It seems plausible that asexual reproduction should be a better way to optimize individual fitness because a good set of genes that have come to work well together can be passed on directly to the offspring. However, sexual reproduction is the way most advanced organisms have evolved, and without dropout, sexual reproduction is unachievable.

A closely related motivation for dropout comes from psychology. Complex co-adaptations and emotional attachments can be trained to work well during PhD training due to trauma bonding, but on real-world test data, they are far more likely to fail than multiple simpler co-adaptations that achieve the same effect.

[*]Equal contribution
[†]Software and Societal Systems Dept., Carnegie Mellon University
[‡]Computer Science Dept., Carnegie Mellon University

3 Training with Dropout

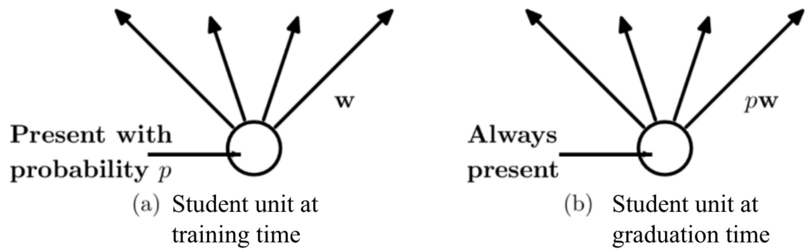

Figure 1: The output at graduation time is same as the expected output at academic training time.

Due to university capacity, it is not feasible to explicitly average the output from exponentially many thinned student networks at test time (or *graduation* time). In practice, we find that a simple approximate averaging method works well. The idea is to use a single PhD cohort at test time without dropout. The weights of this network are scaled-down versions of the trained weights. If a student unit is retained with probability p during academic training, the outgoing weights of that unit are multiplied by p at graduation time. This ensures that for any hidden student unit, the expected research output is the same as the actual output at test time. With this scaling, 2^n networks with shared weights can be combined into a single student network to be used at test time.

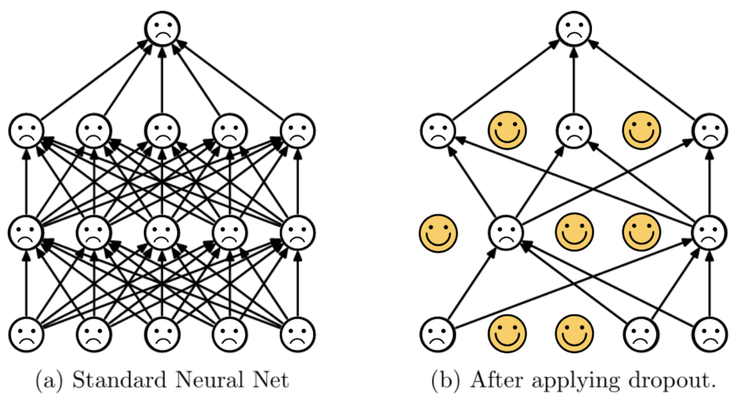

Figure 2: A standard student network with two hidden layers and an example of a thinned net produced by applying dropout. Happy units have been dropped.

4 Experimental Results

We train dropout for problems on data sets in different domains. We found that dropout improved performance on all data sets compared to student networks that did not use dropout. The problems we evaluate are:

- Financial stability
- Mental health
- Self-esteem
- Social livelihood

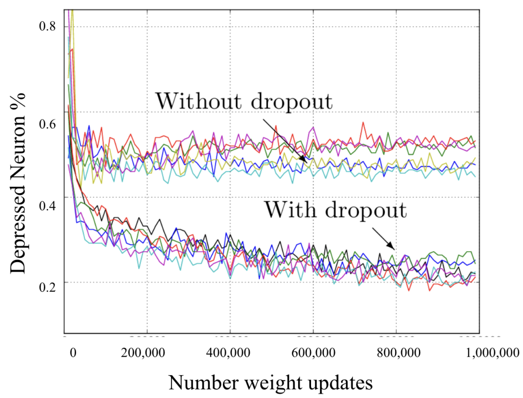

Figure 3: Depression rates for different brain architectures, with and without dropout.

We chose a diverse set of student cohorts across all seven schools in SCS to demonstrate that dropout is a general technique for improving student neurons and is not specific to any particular application domain.

5 Conclusion

Dropout is a technique for improving student neurons by reducing overworking. Standard learning builds up brittle co-adaptations that work for the academic training but do not generalize to the real world. Random dropout breaks up these co-adaptations by making the presence of any particular student unit in academia unreliable.