

Dropout: A Simple Way to Prevent Neural Networks from Overfitting

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov

Journal of Machine Learning Research 15 (2014) 1929-1958

Abstract

Deep neural nets with a large number of parameters are very powerful machine learning systems. However, overfitting is a serious problem in such networks. Large networks are also slow to use, making it difficult to deal with overfitting by combining the predictions of many different large neural nets at test time. Dropout is a technique for addressing this problem. The key idea is to randomly drop units (along with their connections) from the neural network during training. This prevents units from co-adapting too much.

1. Introduction

Deep neural networks contain multiple non-linear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs.

4. Motivation

Overfitting is greatly reduced by randomly omitting half of the feature detectors on each training case. This prevents complex co-adaptations in which a feature detector is only helpful in the context of several other specific feature detectors.

5. Related Work

Dropout can be seen as a way of regularizing a neural network by adding noise to its hidden units.

7.1 Choosing Hidden Unit Dropout Rate

In the simplest case, each unit is retained with a fixed probability p independent of other units, where p can be chosen using a validation set or can simply be set at 0.5, which seems to be close to optimal for a wide range of networks and tasks.

All dropout nets use $p=0.5$ for hidden units and $p=0.8$ for input units.

7.2 Results on MNIST

Table 2 compares different models on MNIST. The MNIST data set consists of 28×28 pixel handwritten digit images. The task is to classify images into 10 digit classes.

Error rates can be further improved to **0.94%** by replacing ReLU units with maxout units.

The lower set of lines also uses 20% dropout for the input layer (i.e., $p=0.8$ for inputs).

8. Conclusion

Dropout improves the performance of neural networks on supervised learning tasks in vision, speech recognition, document classification and computational biology, obtaining state-of-the-art results on many benchmark data sets.