Monitoring MLP's Free Parameters for Generalization

HUNG HAN CHEN
Graphion LLC
8787 Southside Blvd, Suite 503
Jacksonville, FL 32256
UNITED STATES OF AMERICA
hunghchen@mac.com

Abstract: Generalization is one of major concerns for neural network training. In common practice, the number of weights in a MLP network is assumed to be the number of free parameters. This leads to a conclusion: large MLP networks will generalize poorly if their sizes exceed the necessary capacity. However, individual weight in MLP network may not stay as a free parameter since operational condition for weights alters during the course of training. There have been studies showing that larger networks appear to generalize as well as smaller networks, sometimes even better. Therefore this paper constructs a new perspective on MLP's free parameters to address the issue of generalization.

Key-Words: MLP, artificial neural networks, overfitting, generalization, free parameters, automatic pruning

1 Introduction

The ability to be used as an arbitrary function approximation mechanism has made Multi-Layer Perceptrons (MLP), a type of feedforward artificial neural networks, to be utilized extensively for supervised-learning pattern recognition process and a subject of ongoing research in many fields. In most practical use, MLP serves as a nonlinear data-modeling tool by modeling the complex relationship between inputs and outputs.

In applications where the goal is to create a model that generalizes well for unseen data, the issue of overfitting becomes very important. In statistics, overfitting is when free parameters exceed the information content of the data and will lead to overspecified systems that fail to generalize beyond the fitting data. As in common practice, the number of weights in a MLP network is often treated as the number of free parameters. This concept then leads to the conclusion: large MLP networks will generalize poorly if their sizes exceed the necessary capacity.

However, Some studies have shown that this assumption, large MLP networks will generalize poorly, may not be true [1][2]. They offer some simple and intuitive explanations why large networks appear to generalize as well as smaller networks, sometimes even better.

As in a structural model, a free parameter will not be fixed at any particular value by the model hypothesis; On the other hand, individual weight in a MLP network may not stay as a free parameter since operational condition for weights alters during the course of training.

In order to make MLP more practical, it is vital to further investigate how MLP's size matters in generalization. By re-examining the definition of free parameters, this paper presents a new perspective on addressing the issue of overfitting and generalization for MLP networks. If the ability to prevent overfitting by monitoring or controlling the number of free parameters can be realized, the validation step then could possibly be minimized.

The meaning of generalization is discussed in Section 2. The new definition of free parameters for MLP is proposed in Section 3. Section 4 shows how MLP's size matters with 6 simulations on a 4-class segmented image dataset. Automatic pruning is demonstrated in Section 5. Conclusions and further studies are discussed in Section 6.

2 Generalization

Generalization theoretically is a measure that is based upon model performance for unseen data. Since training data can only have finite training samples and may not cover all the possible data, current or future, this measure does make sense. However, what unseen data is and how good generalization is often remain to be defined case by case.

When a model fails with generalization, it often means the model performance on the training data is somewhat satisfied but a disaster on the unseen data. This phenomenon is also known as overfitting. As the unseen data not available for modeling by purpose or by reality, overfitting becomes a major concern for models to be successful or not.