Review of Literature on Growing Neural Networks

Max Ploner
HU Berlin
Science of Intelligence

1 Introduction

This article aims to summarize the existing literature concerned with growing artificial neural networks. For each paper it will list the most significant contribution. The following four questions will guide the summary of each paper:

1. **Why** are models grown? What is the goal or metric the approach is evaluated on?
2. **When** are the models grown?
3. **Where** are the models grown?
4. **How** are the the new parts initialized?

Each paper tries to make progress in answering at least one of the questions. Hence, they can be used to categorize these papers.

2 Reviewed Literature

The following sections give short summaries of each of the papers which were deemed relevant.

2.1 NeST: A Neural Network Synthesis Tool Based on a Grow-and-Prune Paradigm (Dai, Yin, and Jha 2018)

coming soon

2.2 Firefly Neural Architecture Descent: A General Approach for Growing Neural Networks (Wu et al. 2020)

coming soon

2.3 GradMax: Growing Neural Networks Using Gradient Information (Evci et al. 2022)

GradMax focuses on the question **how** new neurons are initialized. They propose initializing new neurons such that the gradient norm of new weights are maximized while maintaining the models function. By enforcing large gradient norms of the new weights, the objective function is guaranteed to decrease in the next step of gradient descent.

When using a step size of $\frac{1}{\beta}$, the loss is upperbounded by:

$$L(W_{new}) \leq L(W) - \frac{\beta}{2}\|\nabla L(W)\|^2$$

The maximum gradient norms are approximated using singular value decomposition (SVD).

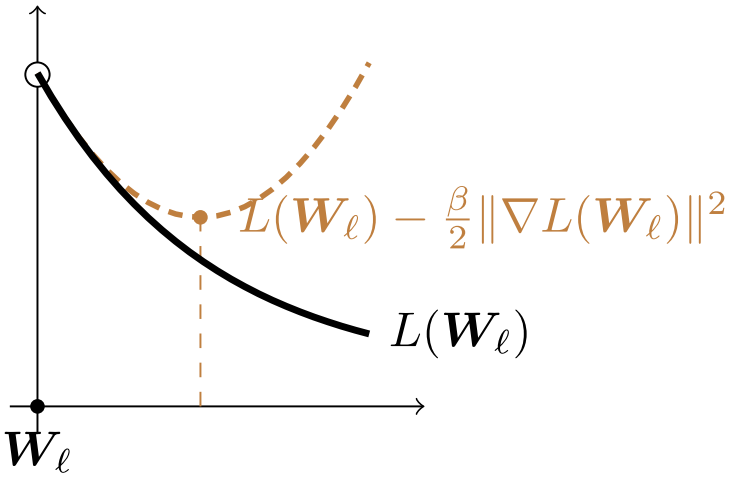

Figure 1: Figure from Evci et al. (2022), showing the described upper bound.

The authors note that this idea could also be utilized to select **where** new neurons should be grown. The decision where to add new neurons could be made by looking at the singular values (e,g, selecting the largest or adding a neuron, once the singular value reaches a threshold). This idea is very similar

to the strategy of Wu et al. (2020) which use a very similar technique to choose **where** to grow neurons (but use a different initialization strategy).

References

Dai, Xiaoliang, Hongxu Yin, and Niraj K. Jha (June 2018). "NeST: A Neural Network Synthesis Tool Based on a Grow-and-Prune Paradigm". In: arXiv:1711.02017. DOI: 10.48550/arXiv.1711.02017. arXiv: 1711.02017 [cs]. URL: http://arxiv.org/abs/1711.02017 (visited on 10/17/2022).

Evci, Utku et al. (Feb. 2022). *GradMax: Growing Neural Networks Using Gradient Information*. DOI: 10.48550/arXiv.2201.05125. arXiv: 2201.05125 [cs]. URL: http://arxiv.org/abs/2201.05125 (visited on 05/28/2022).

Wu, Lemeng et al. (2020). "Firefly Neural Architecture Descent: A General Approach for Growing Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., pp. 22373–22383. URL: https://proceedings.neurips.cc/paper/2020/file/fdbe012e2e11314b96402b32c0df26b7-Paper.pdf.