Analyzing Plasticity Through Utility Scores

Student: Aldas Lenkšas
Supervisors: Wendelin Böhmer, Laurens Engwegen
TU Delft

1. Introduction

- **Continual Learning**: Model keeps getting new data from a new distribution or a new task.
- **Plasticity**: Model's ability to adapt to the changes and keep learning at the same efficiency level.

Symptoms of the *loss of plasticity*: **dead neurons**, **saturated neurons**, **high weight magnitude average**.
To remedy loss of plasticity:

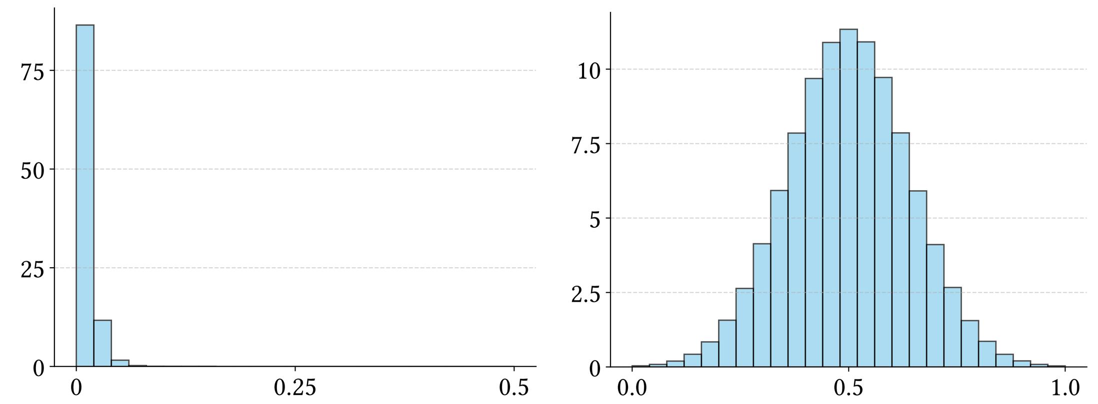
- Standard Machine Learning methods, such as standard **backpropagation** (BP), or **L2 regularization** (L2)
- **Shrink and Perturb** (SnP)
- **Continual Backpropagation** (CBP)
- **CBP variations** with L2 (CBP+L2) or SnP (CBP+SnP)

2. Utility scores

Utility scores evaluate neuron usefulness with the formula by Dohare et al. [1]:

$$\underbrace{\mathbf{u}_l[i]}_{\text{neuron utility}} = \underbrace{\eta \times \mathbf{u}_l[i] + (1-\eta)}_{\text{moving average}} \times \underbrace{|\mathbf{h}_{l,i,t}|}_{\text{output}} \times \underbrace{\sum_{k=1}^{n_{l+1}} |\mathbf{w}_{l,i,k,t}|}_{\text{outgoing weights}}$$

Utility score distribution: Binned histogram with utility scores on the x-axis and their frequencies on the y-axis.



Artificial examples of distributions

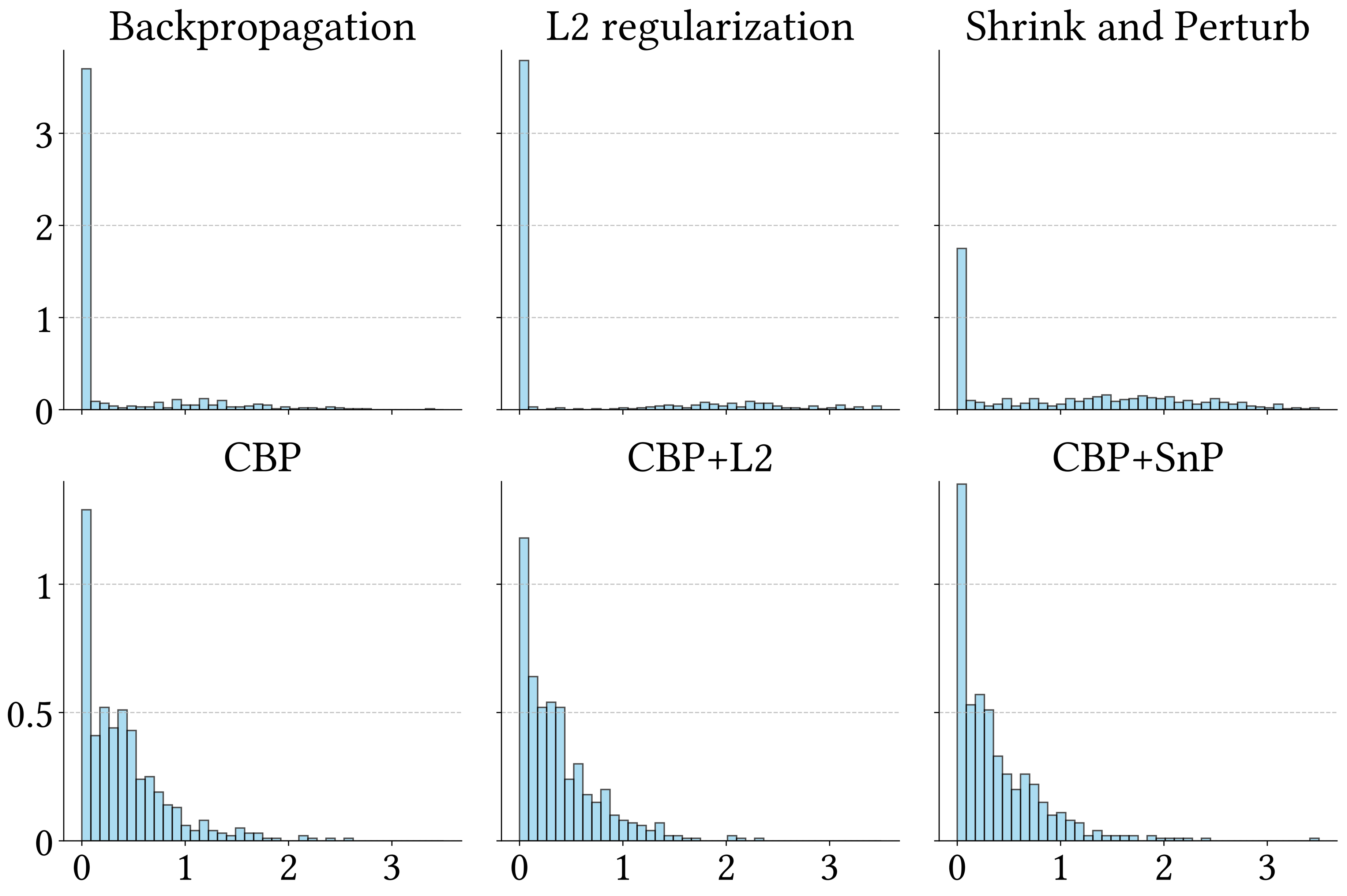
3. Research question

What are the effective utility score distributions for continual learning?

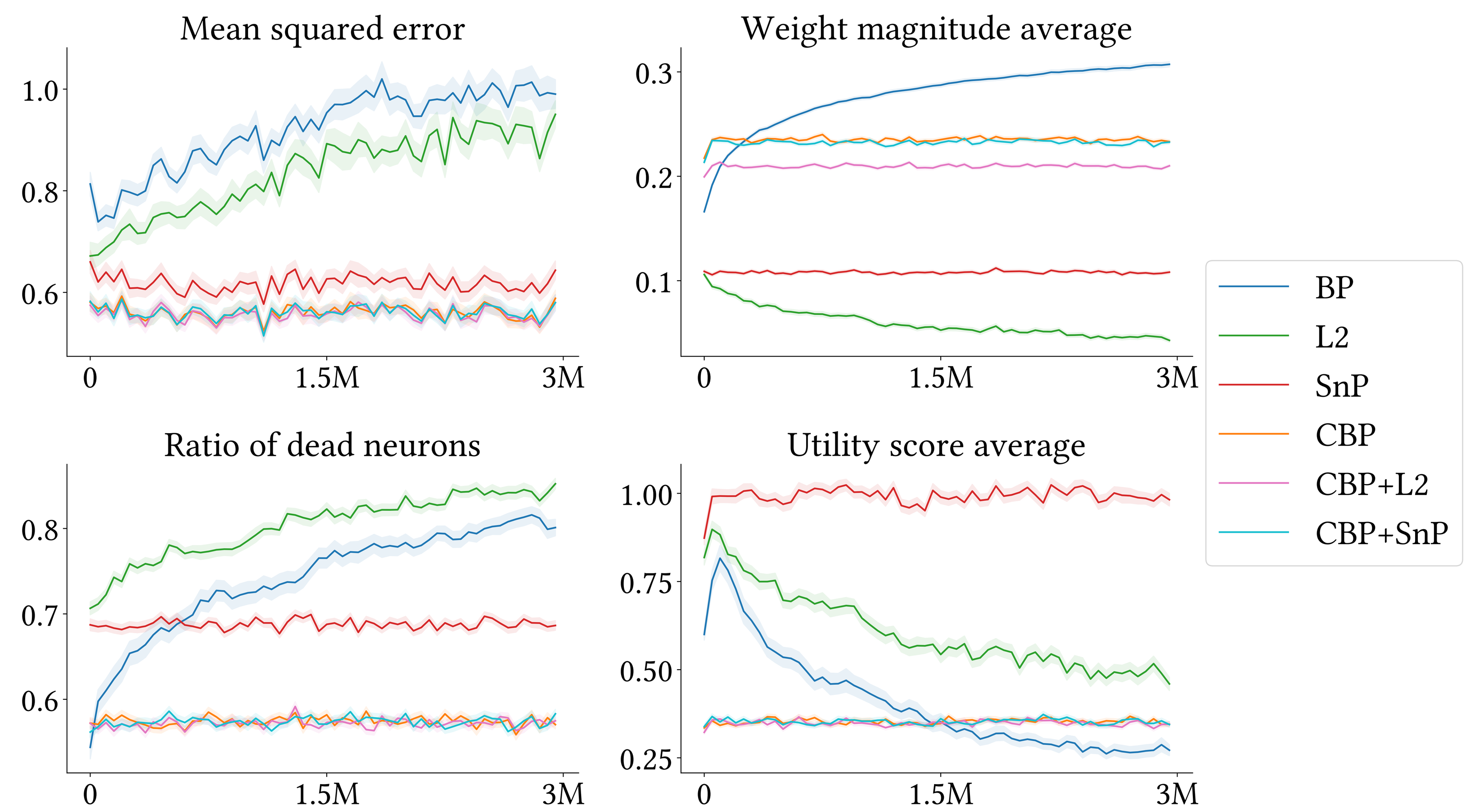
Motivation. Previous research has not looked into utility scores and their relation to the algorithm performance.

4. Slowly-Changing Regression

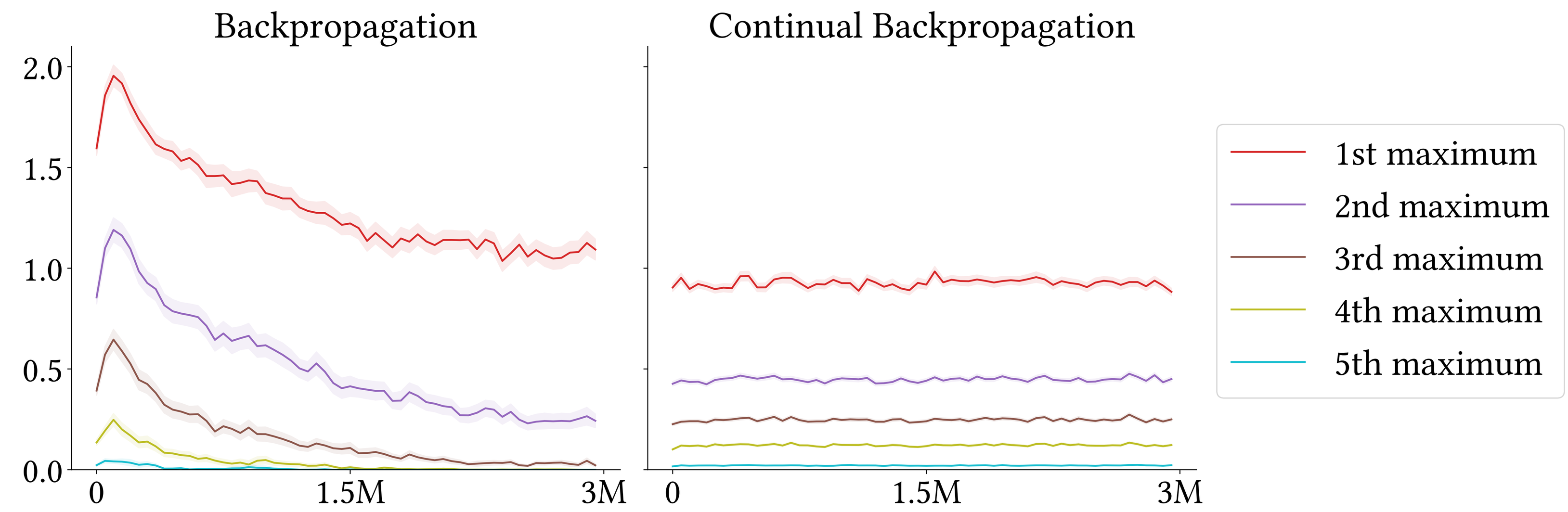
- Input: $\underbrace{x_1 \ldots x_f}_{\text{flipping bits}} \quad \underbrace{x_{f+1} \ldots x_m}_{\text{random bits}}$.
- Outputs are generated by a more complex network.
- Flipping bits flip only every T iterations.



Utility score distributions



Results of different metrics



Individual neuron utility progression

5. Conclusions

- Utility scores and their distributions provide a new angle for analysis of plasticity and an additional explanation for its loss.
- In general, more evenly distributed utilities with low number of scores close to zero correspond to the algorithms that perform better.

6. References

[1] Shibhansh Dohare, J. Fernando Hernandez-Garcia, Qingfeng Lan, Parash Rahman, A. Rupam Mahmood, and Richard S. Sutton. Loss of plasticity in deep continual learning. *Nature*, 632:768–774, 2024. doi: 10.1038/s41586-024-07711-7. Published online: 21 August 2024.