

GASTVORTRAG

Die Arbeitsgruppe Numerical Analysis lädt zu folgendem Vortrag ein:

Steffen Schotthöfer
Karlsruher Institut für Technologie

Low-rank lottery tickets: finding efficient low-rank neural networks via matrix differential equations

Neural networks have achieved tremendous success in a large variety of applications. However, their memory footprint and computational demand can render them impractical in application settings with limited hardware or energy resources. We propose a novel algorithm to find efficient low-rank subnetworks. Remarkably, these subnetworks are determined and adapted already during the training phase and the overall time and memory resources required by both training and evaluating them are significantly reduced. The main idea is to restrict the weight matrices to a low-rank manifold and to update the low-rank factors rather than the full matrix during training.

To derive training updates that are restricted to the prescribed manifold, we employ techniques from dynamic model order reduction for matrix differential equations. Moreover, our method automatically and dynamically adapts the ranks during training to achieve the desired approximation accuracy.

The efficiency of the proposed method is demonstrated through a variety of numerical experiments on fully-connected and convolutional networks.

Zeit: **Montag, 22. August 2022 um 16.15 Uhr**

Ort: **Technikerstr. 13, HSB 6**

Gäste sind herzlich willkommen!

Lukas Einkemmer