Einladung zum Oberseminar Mathematik des Maschinellen Lernens und Angewandte Analysis

Julius-Maximilians-Universität Würzburg
Professur für Mathematik des Maschinellen Lernens

Albert Alcalde Zafra

Friedrich-Alexander-Universität Erlangen-Nürnberg,
Chair for Dynamics, Control, Machine Learning and Numerics

Clustering in pure-attention hardmax transformers

Transformers are extremely successful models in machine learning with poorly understood mathematical properties. In this talk, we rigorously characterize the asymptotic behavior of transformers with hardmax self-attention and normalization sublayers as the number of layers tends to infinity. By viewing such transformers as discrete-time dynamical systems describing the evolution of interacting particles in a Euclidean space, and thanks to a geometric interpretation of the self-attention mechanism based on hyperplane separation, we show that the transformer inputs asymptotically converge to a clustered equilibrium determined by special particles we call leaders. We leverage this theoretical understanding to solve a sentiment analysis problem from language processing using a fully interpretable transformer model, which effectively captures 'context' by clustering meaningless words around leader words carrying the most meaning.

Ort: Mathematik Ost, Seminarraum 40.01.003 Zeit: Mittwoch, 20.11.2024 14:15

Zu diesem Vortrag laden wir Sie herzlich ein.

gez. Leon Bungert