



Life Sciences Seminar

From Loss Landscape Geometry to Weight Recovery in Neural Networks

Flavio Martinelli

EPFL

Host: Tim Vogels

Can we identify the weights of an artificial neural network by probing its input-output mapping? At first glance, this problem seems to have too many solutions because of various symmetries. Yet, we show that the incoming weight vector of each neuron is identifiable up to sign or scaling, depending on the activation function. Our novel approach 'Expand-and-Cluster' can identify layer sizes and weights of a target network for all commonly used activation functions. In my talk, I will give some intuitions on the dynamics of training for a very specific learning setup: a network that regresses the output of another network (teacher-student setup). Then, I will show you how we avoid local minima of gradient descent dynamics and why identification of weights is possible. To conclude, I will present a practical algorithm to perform network identification.

Tuesday, July 30, 2024 01:00pm - 02:00pm

Office Bldg West / Ground floor / Heinzl Seminar Room (I21.EG.101)



This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station.

Please find a schedule of the ISTA Shuttle on our webpage:

<https://ista.ac.at/en/campus/how-to-get-here/> The ISTA Shuttle bus is marked ISTA Shuttle (#142) and has the Institute Logo printed on the side.