Contribution ID: **48** Type: **Invited Talk**

What do neural networks learn? On the interplay between data structure and representation learning

Wednesday, 21 September 2022 14:30 (45 minutes)

Neural networks are powerful feature extractors - but which features do they extract from their data? And how does the structure of the training data shape the representations they learn? We investigate these questions by introducing several synthetic data models, each of which accounts for a salient feature of modern data sets: low intrinsic dimension of images [1], symmetries and non-Gaussian statistics [2], and finally sequence memory [3]. Using tools from statistics and statistical physics, we will show how the learning dynamics and the representations are shaped by the statistical properties of the training data.

[1] Goldt, Mézard, Krzakala, Zdeborová (2020) Physical Review X 10 (4), 041044 [arXiv:1909.11500]
[2] Ingrosso & Goldt (2022) [arXiv:2202.00565]
[3] Seif, Loos, Tucci, Roldán, Goldt [arXiv:2205.14683]

Presenter: Prof. GOLDT, Sebastian (International School of Advanced Studies (SISSA))

Session Classification: Understanding Machine Learning