Contribution ID: **21**

Type: **not specified**

Injectivity of ReLU networks: perspectives from integral geometry and statistical physics

Wednesday 28 September 2022 10:00 (20 minutes)

We consider the well-posedness of inferring the input of a randomly-initialized large ReLU neural network from its output, i.e. characterizing injectivity.
Focusing on layerwise injectivity properties, we discuss recent work connecting this question to spherical integral geometry, and present a conjecture
for a sharp injectivity threshold (in terms of the expansivity of the layer) based on a transition in the expected Euler characteristic of a particular random set.
Showing that injectivity is also equivalent to a property of the ground state of a spherical perceptron in statistical physics, we then leverage the non-rigorous
replica symmetry breaking theory to obtain analytical equations satisfied by the injectivity threshold.
Efficiently solving the zero-temperature full replica symmetry breaking equations yields a conjectured threshold at odds with the integral geometry approach described above.
Finally, using a classical approach based on Gordon's min-max theorem, we show that the replica symmetric calculation, although non-exact, can already disprove the Euler characteristic threshold, leaving open to understand the discrepancy between these predictions.

Presenter: MAILLARD, Antoine (ETH Zurich)