Deep learning has been immensely successful at a variety of tasks, ranging from classification to artificial intelligence. Yet why it works is unclear. Learning corresponds to fitting training data, which is implemented by descending a very high-dimensional loss function. Two central questions are (i) since the loss is a priori not convex, why doesn't this descent get stuck in poor minima, leading to bad performance? (ii) Deep learning works in a regime where the number of parameters can be larger, even much larger, than the data to fit. Why does it lead to very predictive models then, instead of overfitting?

Here I will discuss an unexpected analogy between the loss landscape in deep learning and the energy landscape of repulsive ellipses, that supports an explanation for (i). If times permit I will discuss (ii), more specifically the surprising finding that predictive power continuously improves by adding more parameters.

Tuesday December 18th 2018 at 12:00

Auditoire CUB II, room BSP 234, (Cubotron), EPFL