Phase transitions in Artificial Inteligence

Gonçalo Oliveira

Instituto Superior Técnico

IST, Março 2024

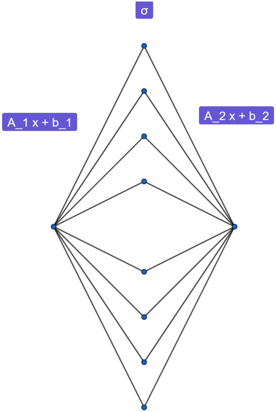
Phase transitions (examples)

“Sudden changes of behavior”: When a relevant function has a **discontinuity** in one of its derivatives.

- Gas-liquid.
- Ferromagnetism.
- Message/disease propagating in a network (or graph).

Examples from several areas of science/engineering, including AI.

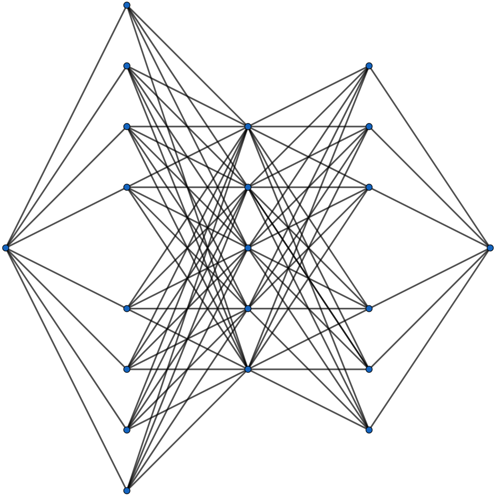
- Neural network $\Rightarrow$ function $f : \mathbb{R}^{N_{in}} \to \mathbb{R}^{N_{out}}$, which to an input associates an output. Schematically, for one layer



- For $b_1 \in \mathbb{R}^{N_{in}}$, $b_2 \in \mathbb{R}^{N_{out}}$ and $A_1 \in M_{N \times N_{N_{in}}}$, $A_2 \in M_{N_{out} \times N}$

$$f(x) = A_2 \sigma(A_1 x + b_1) + b_2.$$

- Neural network $\Rightarrow$ function $f : \mathbb{R}^{N_{in}} \rightarrow \mathbb{R}^{N_{out}}$. An example with three layers:



$$f(x) = A_4\sigma(A_3\sigma(A_2\sigma(A_1 x + b_1) + b_2) + b_3) + b_4.$$

- Neural network $\Rightarrow$ function $f : \mathbb{R}^{N_{in}} \to \mathbb{R}^{N_{out}}$. For one layer

$$f(x) = A_2\sigma(A_1 x + b_1) + b_2,$$

with

$$\sigma : \mathbb{R}^N \to \mathbb{R}^N$$

a fixed function (activation function).

- The network can be taught pairs $\{(x_i, y_i)\}_{i\in I}$: Minimize

$$\mathcal{L}(\mathbf{A}, \mathbf{b}) := \sum_{i\in I} (f(x_i) - y_i)^2\,,$$

with $\mathbf{A} = (A_1, A_2)$ and $\mathbf{b} = (b_1, b_2)$.

- Evolve $\mathbf{A}(t), \mathbf{b}(t)$ with $-\nabla\mathcal{L}$. Then, the values $f_t(x_i)$ evolve according to:[1]

$$\frac{d}{dt}(f_t(x_i) - y_i) = -\sum_{j\in I} \Theta_{ij}(f_t(x_j) - y_j),$$

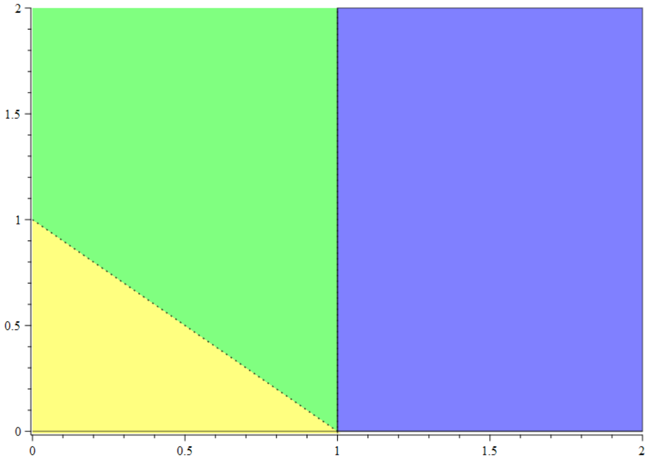
for a matrix Θ known as Neural Tangent Kernel (NTK).

[1] When $N \to +\infty$

Spectral gap and spectral ratio

- From the equation
$$\frac{d}{dt}(f_t(X) - Y) = -\Theta(f_t(X) - Y),$$
we find
$$f_t(X) - Y = e^{-\Theta t}(f_0(X) - Y).$$
- If
$$\lambda_{min} := \text{smaller eigenvalue } \Theta,$$
we obtain
$$|f_t(X) - Y| \lesssim e^{-\lambda_{min} t}.$$
$\implies \lambda_{min} \sim$ learning rate!
- If
$$\lambda_{max} := \text{largest eigenvalue } \Theta.$$
$\implies \frac{\lambda_{max}}{\lambda_{min}} \sim$ stability of learning.
- Varying the network parameters ... phase transitions in λ_{min} e $\frac{\lambda_{max}}{\lambda_{min}}$.

Phase diagram for $\operatorname{tr}\Theta$ (L. Carvalho, J. Costa, J. Mourão, O.)



Phase diagram for the behavior of $\operatorname{tr}\Theta$ in wide and deep networks.

To do:

- Find relevant quantities to study.
- Learn techniques which can be used to establish the existence of phase transitions (Crandall–Rabinowitz theorem, Lyapunov–Schmidt reduction).
- Prove the existence of phase transitions in the problem at hand.
- Understand what is the best phase (for our goals).

Thank you!