Contribution ID: **75**   Type: **not specified**

Neural Network Pruning: from over-parametrized to under-parametrized networks

Wednesday 21 October 2020 17:00 (30 minutes)

This talk will provide an introduction to the concept of over-parametrization in neural networks and the associated benefits that have been identified from the theoretical and empirical standpoints. It will then present the practice of pruning as both a practical engineering intervention to reduce model size and a scientific tool to investigate the behavior and trainability of compressed models under the "lottery ticket hypothesis" (Frankle and Carbin, 2018). Finally, it will demonstrate features for tensor and neural network pruning in PyTorch.

Presenter: Dr PAGANINI, Michela (Facebook AI Research)

Session Classification: Plenary