Phase transitions in mini-batch size for sparse and dense deep neural networks

Raffaele Marino
University Of Florence, Florence, Italy

The widespread diffusion of neural networks in many scientific fields urges for a better understanding of the processes that underlie their training. The discipline that studies how well simple devices like Turing machines or more abstract constructions, such as classifier systems, can learn and infer from observed data after the training process, is the statistical learning theory. The statistical learning theory is at the cornerstone of Machine Learning, and it deals with the statistical inference problem of finding a predictive function based on data. Even though neural networks are well analyzed from statistical learning, they have become an active subfield of research in statistical mechanics. In this area a major role is played by phase transitions that regulates what is achievable in principle (information theoretical thresholds) and what is achievable in practice (algorithmic thresholds). Such a vast field of possible applications has been found for networks, as a huge variety of systems can be described in terms of interconnected elements. Recently, statistical mechanics tools have been also applied in the realm of artificial intelligence, for building up consistent theories of deep learning. In simple words, deep learning can be seen as a fully connected neural network that takes some data, composed by input and targets, and learns the rules for forecasting new input data. Over the last decades the practitioners of neural networks have developed many very useful tricks and smart procedures, like mini-batch, dropout and others several regularizations for speeding up the training steps. A theory justifying many of these choices is often lacking, and so it is very difficult to make optimal choices for who is not an expert. Among these "tricks" the use of the so called mini-batch, introduced as a technical requirement for dealing with huge databases, actually turns out to be crucial for the optimal training. In machine learning, a mini-batch is a subset of the full dataset that is used to train a model. Rather than training the model on the entire dataset at once, the training data is divided into smaller batches, or mini-batches, which are fed to the model one at a time. The size of the mini-batch is a hyperparameter that can be tuned to optimize the training process. A larger mini-batch size can lead to faster training times, but it can also make it harder for the model to converge to a good solution. A smaller mini-batch size can improve the model's convergence but may slow down the training process. In this work, we present the existence of phase transitions in sparse and dense neural networks, in the realm of the teacher-student scenario of deep learning. We show that the mini-batch size m plays a fundamental role in inferring the weights of the teacher neural networks. More precisely, we show that above a certain value of m, named mc, the inference is always possible, while below it the inference is impossible. This phenomenon seems to be independent of the architecture of the neural network used.