Seminar Review 4.20-4.27

Jiaxin Hu

04/20/2020

IFDS 4.20

Title: Bad global minima can be reached by overparameterized model with SGD

Author: Dimitris Papailiopoulos

This talk explains why overparameterized model trained with SGD can work well in many cases. Contrast to the consensus that "there is no bad global minima", the overparameterized SGD is attracted by the bad global minimia. The bad global minima can be considered as a complex boundary in classification and it can lead vanilla methods to get 100% training accuracy while poor performance in tests. The initial point can be set randomly with unlabeled data, which leads the SGD converge to the bad minima quickly. To eliminate the effect from the random initialization, a regularization term can be added to let SGD escape and have a good test performance finally.

Questions:

1. As presenter shows, the vanilla SGD with random initialization and SOTA(the overparameterized SGD method) with adversarial initialization point have similar performance in practice. Which one should we use? Is there any pre-judgement steps?

Possible Answer: The overparameterized SGD has smaller computational complexity. The data which has poor performance with traditional methods may suitable for this method.