Hyomin Seo
Professor Curro
November 2020
2020 ECE472

QUIZ 8

1. Make an argument about super-convergence as it relates to epoch-wise double descent.

Reading Links :

https://arxiv.org/pdf/1708.07120.pdf : Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates

https://arxiv.org/pdf/1912.02292.pdf: Deep Double Descent: Where Bigger Models and More Data Hurt

Super-convergence in neural nets is, simply put, the use of large learning rates that regularizes the network, requiring a reduction of all other regularization forms to preserve a balance between underfitting and overfitting. The inductive biases determine the performance of stochastic gradient descent, where with the appropriate inductive biases the SGC chooses its model of perfect training accuracy, among a limited set of models. This results in complete dismissal of SGD inductive biases.

The deep-double descent phenomenon occurs when the model performance improves, then shortly gets worse as the model begins to overfit with increased test error due to bias-variance tradeoff; the, finally the model improves to approximately zero training error with increased model/ data size or training time. As the title 'double descent' says, the model goes through 2 phases to eventually become a nearly perfect model. Such phenomenon has its implication on model complexity, optimizing quantity of data and training time.

The Epoch-wise double descent is when the training time increases while the model is fixed. The performance initially follows the U curve; when the sample's quantity exceeds the model complexity, when the model passes the stage, when significant model complexity finally exceeds the sample quantity, then mitigating the training time. More simply put, with fixed model size, the test and train error goes through these stages: decreasing, increasing, then dropping, according to change on training time. The highest test error appears when the model quantity just fit the train set.

Implementing Super-convergence when epoch-wise double descent phenomenon is detected might not be as effective as when used separately. The benefit of each method lies in their quantity of iteration - Super-convergence's merit is that it generally requires fewer iteration, and epoch-wise double descent's value is with more frequent iteration, the model gets significantly greater accuracy. However, if the super-convergence were to be adopted on epoch-wise double descent, it will result in losing either of the merits, increasing iteration for super-convergence, and eliminating great accuracy by cutting off epoch-wish double descent during its progress. Therefore, the overall idea of adopting super-convergence in relation to epoch-wish double descent will not yield the double of each method's value, but only the elimination of both.