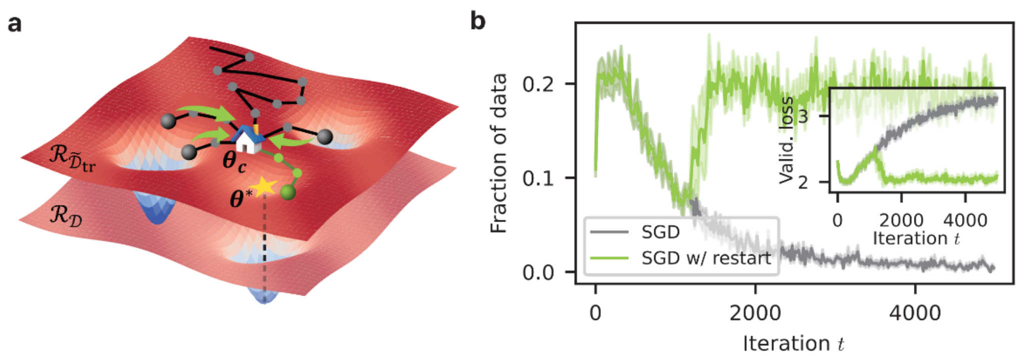
Mitigating Overfitting in Neural Networks with Stochastic Restarting under Noisy Labels

Youngkyoung Bae[1*], Yeongwoo Song[2*], and Hawoong Jeong[2,3]
[1]*Department of Physics and Astronomy, Seoul National University*
[2]*Department of Physics, Korea Advanced Institute of Science and Technology*
[3]*Center for Complex Systems, Korea Advanced Institute of Science and Technology*
**Equal Contribution*

Capitalizing on the theoretical success of stochastic restarting, numerous algorithms incorporating the restarting strategy have begun to emerge in diverse fields. In this work, we demonstrate that restarting from a checkpoint can significantly improve generalization performance when training deep neural networks (DNNs) with noisy labels. In the presence of noisy labels, DNNs initially learn the general patterns of the data but then gradually overfit to the noisy labels. To combat this overfitting phenomenon, we developed a method based on stochastic restarting, which has been actively explored in the statistical physics field for finding targets efficiently. By approximating the dynamics of stochastic gradient descent into Langevin dynamics, we show that restarting can provide great improvements as the batch size and the proportion of corrupted data increase. We then empirically validate our conjecture, confirming the significant improvements achieved by restarting. An important aspect of our method is its ease of implementation and compatibility with other methods, while still yielding notably improved performance. We envision it as a valuable tool that can complement existing methods for handling noisy labels.



(a) Schematic of stochastic gradient descent (SGD) dynamics with stochastic restarting. The network parameters vector $\boldsymbol{\theta}$ evolves via SGD to find an optimal value $\boldsymbol{\theta}^*$ on the training risk landscape $\mathcal{R}_{\widetilde{\mathcal{D}}_{\text{tr}}}$ (upper colormap), which differs from the true risk landscape $\mathcal{R}_{\mathcal{D}}$ (lower colormap) due to corrupted data. Here, $\boldsymbol{\theta}$ resets to the checkpoint $\boldsymbol{\theta_c}$ (home icon) with the restart probability r and restarts from $\boldsymbol{\theta_c}$. (b) Fraction of correctly predicted data with wrong labels during training with SGD (gray) and SGD with restart (green). The inset shows the validation losses during training.