

We show that any model trained by a stochastic gradient method with few iterations has vanishing generalization error. We prove this by showing the method is algorithmically stable in the sense of Bousquet and Elisseeff. Our analysis only employs elementary tools from convex and continuous optimization. Our results apply to both convex and non-convex optimization under standard Lipschitz and smoothness assumptions. Applying our results to the convex case, we provide new explanations for why multiple epochs of stochastic gradient descent generalize well in practice. In the nonconvex case, we provide a new interpretation of common practices in neural networks, and provide a formal rationale for stability-promoting mechanisms in training large, deep models. Conceptually, our findings underscore the importance of reducing training time beyond its obvious benefit.