Key contributions

The paper's key findings "problematize the traditional view of generalization by showing it is incapable of distinguishing between different neural networks that have radically different generalization performance." The results of the experiments can be summarized in that deep neural networks easily fit random labels. The paper further shows that these results rule out many of the possible explanations for generalization performance of state-of-the-art neural networks. They provide a notion of the effective capacity of machine learning models.

Strengths

The paper presents thorough experiments and results that show that the training process for multiple standard architectures is unaffected by whether there is a relationship between instances and the class labels. It shows that their findings imply deep neural networks are able to fit perfectly any set of labels. They show that the effective capacity of several successful architectures is large enough to "shatter the training data."

Weaknesses

Even though the paper provides experimentation, results, and theory on the traditional view of generalization, it did not reach a solution or approach to tell apart models that generalize well. The paper provides an experimental framework for the effective capacity of machine learning models, but it was unable to provide a scientific way/method of measuring such effective capacity. Ultimately, it does not provide a concrete answer to the question it proposes in the beginning: "what distinguishes neural networks that generalize well from those that don't?"

My Takeaways

The paper’s experiments and findings that explicit forms of regularization do not adequately explain the generalization error of neural nets resonated with some of my prior experience when looking for hyperparameters. I found that regularization did not have too much of an impact on my small model. The paper also makes me think how the difference between models that generalize vs ones that do not (our incomplete understanding of it) could be related to the fact that neural nets are uninterpretable by us. If we could accomplish one or the other, we may be able to learn more about the other.