Generalizable Scientific Machine Learning

Vivak Patel[1][2]

In typical consumer machine learning applications, such as facial detection in images and automatic language translation, generalization to unobserved examples has been disruptively successful. However, as machine learning enters scientific and engineering applications, equally successful generalization has remained elusive. For example, machine object detection techniques used in self-driving vehicles can be easily fooled by changes to road signs that would fail to fool human drivers [Athalye and Sutskever, 2017]. Using similar techniques, machine learning tools for the analysis and control of distributed energy resources in power systems could also be easily fooled by modified sensor data, which could lead to expensive reliability issues [Young et al., 2016]. Importantly, while these example demonstrates the challenges of generalization in scientific machine learning, they underscore that the notion of generalization is fundamentally different in scientific machine learning in comparison to consumer machine learning.

Generalization is fundamentally different in scientific machine learning.

Generalization in scientific machine learning cannot be defined simply as successful operation on unobserved examples. In addition to this requirement, generalization in scientific machine learning must also account for those demands unique to scientific and engineering problems: adversarial defensibility; scientifically sensible and interpretable outputs; and quantitative bounds of uncertainty arising from observable and unobservable random sources.

Adversarial Defenses. An adversarial example is a contrived input to a machine learning methodology that is sufficiently similar to a real example such that it fails to fool humans, but results in dramatic prediction errors by the machine learning methodology. Importantly, adversarial examples are easily generated by solving simple optimization problems without any knowledge of the inner workings of the machine learning technique. In the case of a machine-learning-enabled smart grid, adversarial examples can be implemented by modifying just handful of sensors and observing the response of the grid. Thus, if machine learning methods are not protected against these examples, adversarial examples can result in fundamentally misleading results or potentially costly errors in scientific and engineering problems. □

Scientifically Sensible Outputs. If a scientific or engineering principle dictates that an increase in one variable increases some output, then a rigorous machine learning methodology must enforce this principle. However, as observed in a consumer machine learning application to estimating housing prices [Bonakdarpour, 2017], such outcomes are not guaranteed. Thus, in such scientific and engineering applications as machine-learning-based control of a smart grid, the loss of scientifically sensible machine learning predictions naturally limits the credibility and utility of such techniques. □

Quantitative Uncertainty Bounds. As it currently stands in consumer machine learning techniques, the uncertainty of any output is given by *qualitative* bounds that rely on the rarely-valid assumption of independent and identically distributed examples. Moreover, the sources of uncertainty that drive the random variations in observations are often ignored. However, as understood from statistics, an estimate's or prediction's value is either bolstered or tempered by its estimated uncertainty. Moreover, as is well understood in applied mathematics, the uncertainty measurement of a prediction must also be interpreted by the sources of uncertainty, which include such observable sources of randomness as measurement error; and such unobservable sources of randomness as stochastic energy demands in power systems and stochastic consumption in inventory and manufacturing control. Thus, for scientific machine learning, quantitative uncertainty bounds must be propagated in a principled manner to ensure an appropriate interpretation. □

To reiterate, generalization for scientific machine learning is not only defined as successful operation on unseen observations—as it is for consumer machine learning—but it also requires adversarial defenses, scientifically sensible outputs and quantitative uncertainty bounds. Given this more challenging notion of scientific machine learning generalization, how do we achieve it?

[1] Department of Statistics, University of Chicago
[2] Mathematics and Computer Science, Argonne National Laboratory

For generalization, bridge stochastic optimization, applied mathematics and probability.

The typical notion of generalization—the successful operation on unseen observations—is hypothesized to be the consequence of finding flat minimizers of the empirical risk objective function during training [Keskar et al., 2016]. In this same work, flat minimizers were shown to be preferentially selected by highly stochastic optimization techniques.[3] This observation was exploited by combining stochastic optimization and applied probability sampling to develop training objective functions and stochastic optimizers that preferentially selected these flat minimizers [Chaudhari et al., 2016]. Thus, improved consumer machine learning generalization was achieved by bridging techniques in stochastic optimization and probability.

Similarly, rigorously bridging stochastic optimization, areas of applied mathematics and probability holds the promise to address the challenges of generalization for scientific machine learning. Importantly, stating what these bridges might be is rather simple, but rigorously understanding their implication and overcoming the computational challenges that they impose will be a challenging and extensive endeavor. We give examples below.

Sensitivity Analysis for Adversarial Defenses. Adversarial examples are generated by identifying small perturbations of the input that generate a sufficiently large deviation in the output. An applied mathematician would recognize this as identifying the sensitivity of the output relative to the input. Then, from heuristic reasoning, we should be able to protect machine learning methods against adversarial examples by adding a penalization term on sensitivities during training. However, can we show that this approach will lead to adversarial defense? Are there implications for uncertainty quantification? Importantly, how do we flexibly implement the tools for the unique computational challenges in machine learning? □

Chance Constraints for Scientifically Sensible Outputs. Many scientific principles can be expressed as inequality relationships between outputs. Therefore, we can readily express scientific principles using inequality constraints imposed on the risk. However, such constraints would inherently be random. Given these chance constraints, can we show that a solution to the risk minimization problem exists? If so, what are the properties of such a solution and how do we practically integrate these constraints during the stochastic optimization? If not, can we relax the constraints and what are the implications for generalization? □

Probabilistic Analysis for Uncertainty Bounds. From classical asymptotic statistics, quantitative uncertainty analysis of estimated machine learning model requires determining the local geometry of the risk function, yet the necessary local geometric information is hard to extract because of the complexity of the models and the sheer volume of data. Fortunately, because of the exploratory nature of stochastic optimizers, the local geometry can be extracted by aggregating information from the typically, independent stochastic gradient steps. However, such an approach requires understanding complex stochastic processes generated by the optimizers. Given this requirement, what is an optimal estimator of the local geometry from the information generated by the stochastic optimizers? What is a practical, computable estimator of the local geometry from the information generated by the stochastic optimizer? Can this information be used to improve stochastic optimizers? How can we propagate the model uncertainty bounds extracted from the local geometry into predictions? □

References

Anish Athalye and Ilya Sutskever. Synthesizing robust adversarial examples. *arXiv preprint arXiv:1707.07397*, 2017.

Mahtiyar Bonakdarpour. Shape-constrained random forests and discrete choice. 2017.

Pratik Chaudhari, Anna Choromanska, Stefano Soatto, and Yann LeCun. Entropy-sgd: Biasing gradient descent into wide valleys. *arXiv preprint arXiv:1611.01838*, 2016.

Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*, 2016.

Vivak Patel. The impact of local geometry and batch size on the convergence and divergence of stochastic gradient descent. *arXiv preprint arXiv:1709.04718*, 2017.

Rob Young, John McCue, and Christian Grant. The power is on: How iot technology is driving energy innovation. 2016. URL `https://goo.gl/TQAXs7`.

[3] We showed that the mechanism behind this is analogous to classical results on the convergence and divergence of gradient descent. Thus, even deterministic methods can be forced into flat minimizers, but stochastic methods will do so more naturally and more inexpensively [Patel, 2017].