NJIT | INSTITUTE FOR DATA SCIENCE

Department of Data Science

Data Science Seminar Series

Small changes, big effects? Quantifying the sensitivity of machine learning analyses

Soumya Ghosh, Ph.D.
Research Scientist
MIT-IBM Watson AI Lab – IBM Research

Date: Wednesday, March 2nd, 2022
Time: 11:00 AM – 12:00 PM EST
Location: Zoom Virtual Room
Web Link: Zoom Meeting Room Link

This is an exciting time for machine learning research. Machine learning advances have percolated to both the industry and other academic disciplines, including the social and physical sciences, healthcare, and medicine. However, with increased adoption comes the potential for increased societal impact and harm. Like any other approach that learns from data, inferences drawn from a machine learning analysis are contingent upon the available data and the modeling and algorithmic assumptions made during the analysis. Quantifying the sensitivity of the results of such an analysis to fluctuations in the training data, perturbations in modeling, and algorithmic assumptions is essential for trusting the results.

In this talk, I will describe my recent work on computationally efficient methods for quantifying sensitivity and assessing the robustness of machine learning analyses. First, focusing on Gaussian processes (GP), I will describe approaches for probing their robustness to a key modeling assumption — the choice of the kernel function. I will show how the resulting method provides a model criticism tool for practitioners to flag robustness issues in their GP-based analysis. Next, I will describe methods for quantifying the sensitivity of structured latent variable models to natural fluctuations in the data. I will show how to exploit ideas based on the infinitesimal jackknife to develop fast and accurate methods to approximate the effect of dropping or otherwise perturbing a subset of the training data. I will describe how the resulting approximations enable sensitivity checks that would have been computationally prohibitive otherwise. Finally, I will end with future research directions, including exciting intersections with science and healthcare

Soumya Ghosh is a Research Scientist at the MIT-IBM Watson AI Lab within IBM Research. He completed his M.S. in computer science from the University of Colorado, Boulder, his Ph.D. from Brown University, and was a postdoctoral researcher at Disney research. He is broadly interested in statistical machine learning and its applications in science and healthcare. His recent research has focused on developing procedures for quantifying the sensitivity of inferences drawn from machine learning analyses, on approaches for quantifying uncertainties in neural network predictions, and on combining the complementary strengths of probabilistic graphical models and flexible neural networks to model the progression of neurodegenerative diseases. He has been the recipient of an outstanding technical achievement award at IBM Research (2021), outstanding reviewer awards from CVPR (2019, 2021), a Brown University graduate fellowship (2009-10), and a JN Tata endowment scholarship (2004-05).