Normalizing flows for density estimation and uncertainty quantification in Earth data

Machine learning (ML) is a powerful tool that is commonly applied to geophysical problems. A major limitation of ML models, however, is that they only provide a point estimate for the predicted value and fail to provide uncertainty associated with that prediction. Normalizing flows, which are a class of probabilistic generative models, present a solution to this problem by allowing ML algorithms to accurately predict as well as approximate the underlying data distribution in order to provide both density estimation and uncertainty metrics on model outputs. Normalizing flows rely on transformations from a simple probability distribution (e.g. a standard normal distribution) into a more complex distribution (e.g. data distribution) through a series of tractable, invertible, and differentiable functions. The unknown density can then be approximated and sampled by sampling from the initial, simple probability distribution and applying the trained transformations mentioned above. Geophysical datasets have several qualities that would make density estimation and uncertainty quantification using normalizing flows effective: 1) they are quantitative, continuous measurements, 2) they are often high-dimensional, and 3) they exhibit high amounts of aleatoric uncertainty (i.e uncertainty due to natural stochasticity). Normalizing flows could be useful in not only providing uncertainty measurements for computer-aided geophysical interpretation, but also for understanding the underlying probability distributions governing complex earth systems. This study is a proof of concept showcasing preliminary results of applying normalizing flows to unknown distributions. Preliminary results show that by probabilistically mapping the underlying distributions, these models can both model uncertainty and generate realistic synthetic data. In the future, these techniques will be used to model earth-related phenomena where uncertainty plays a key role, such as in climate modeling, P- and S-wave arrival time identification, and task-specific signal processing and compression.

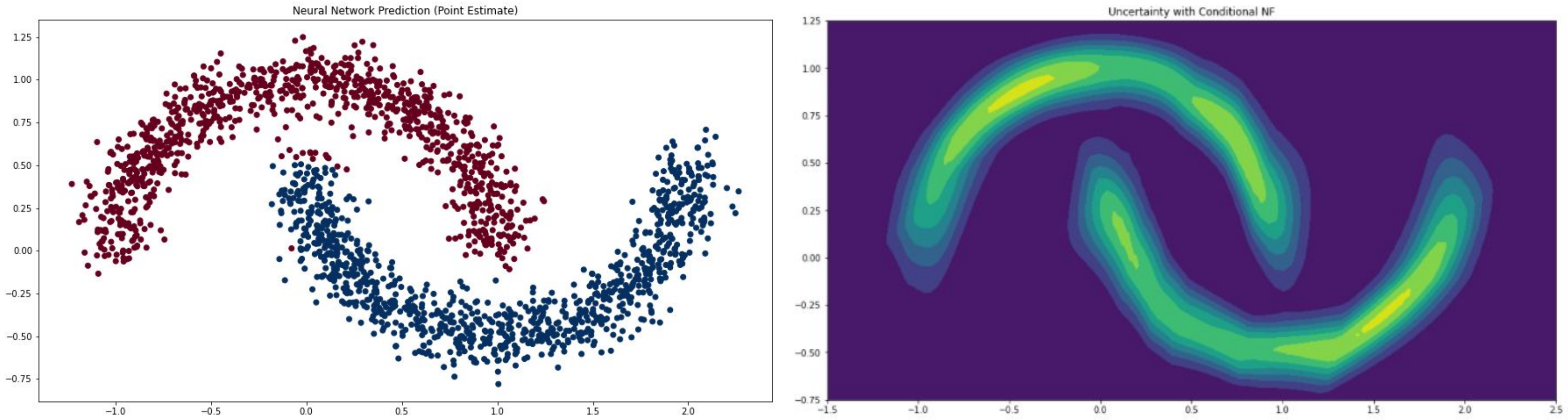

Figure 1. Neural network predictions on scikit-learn moons dataset using standard dense neural network (left) and normalizing flow (right). While the dense neural network performs well with minimal training, point estimates provide minimal information into the uncertainty of each prediction. The normalizing flow accurately classified both moons while quantifying uncertainty using log probabilities.