SAND2020-4999C

Sandia National Laboratories

Teaching Neural Networks to Say '*I Don't Know*'

Comparing Non-Bayesian Out of Distribution Data Detection Methods

Phillip Lewis (6362), Tim Draelos (6362), Sarah Ackerman (6362)

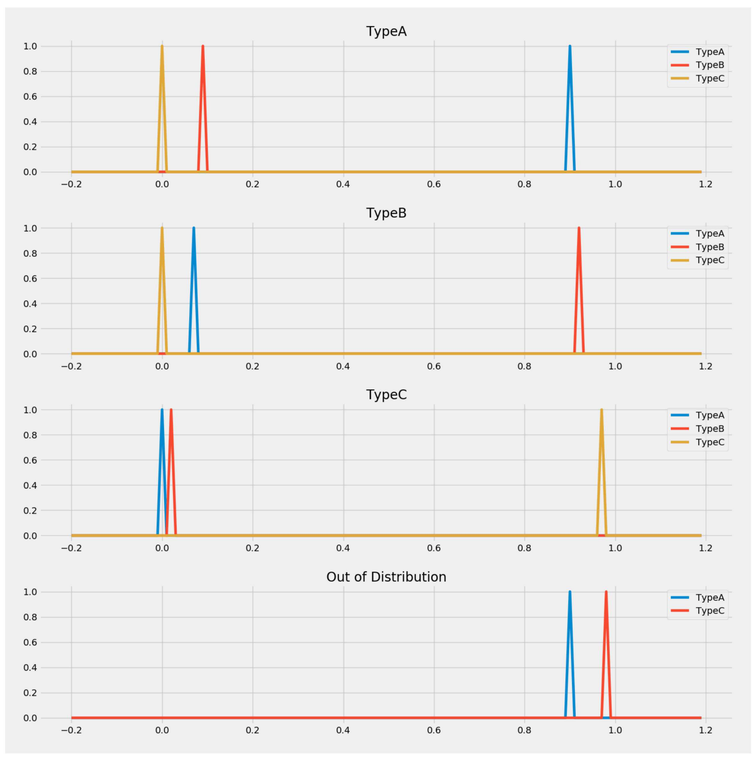
Motivation

- Deep neural networks with softmax outputs make overconfident predictions, particularly for out of distribution data.
- As part of an Automatic Classification, Detection, and Geolocation (ACDG) project deliverable, non-Bayesian alternatives were compared using a proxy 1D dataset.

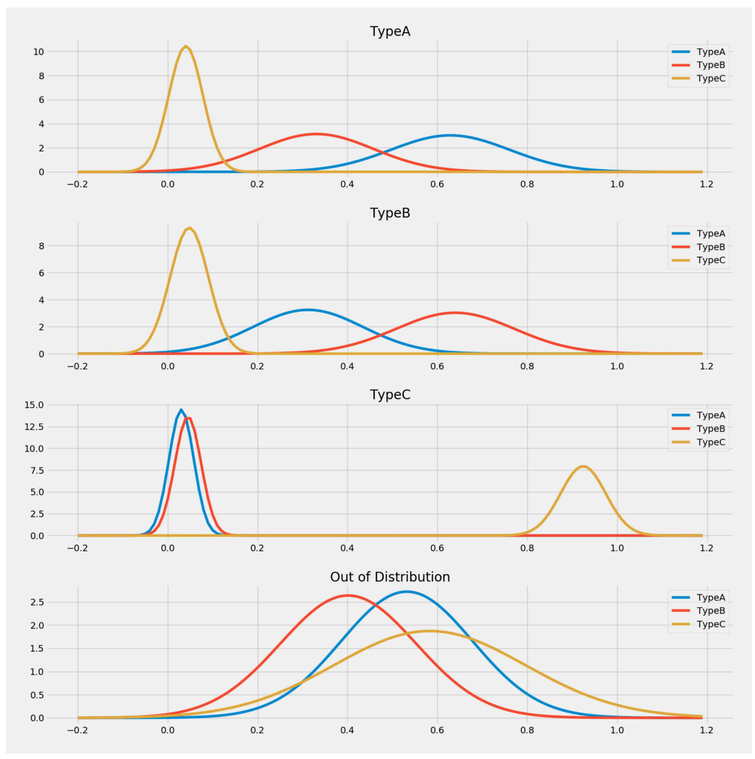
Approach

- Trained and tested on three classes. Tested against an additional out of distribution class.
- Compared baseline convolutional model, an adaptive dropout method (concrete dropout), and a Dirichlet based loss method (evidentiary deep learning).

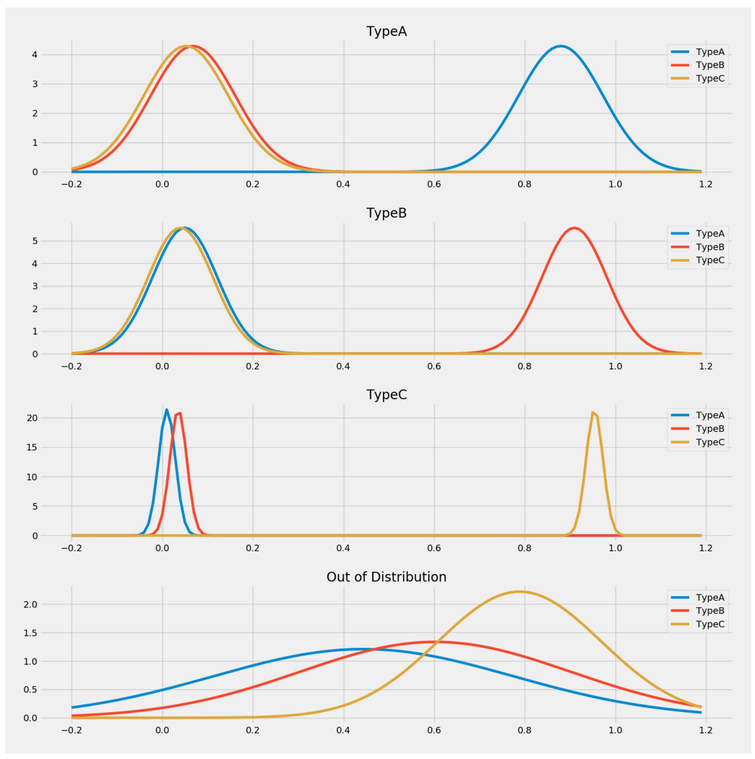
Baseline Softmax



Concrete Dropout

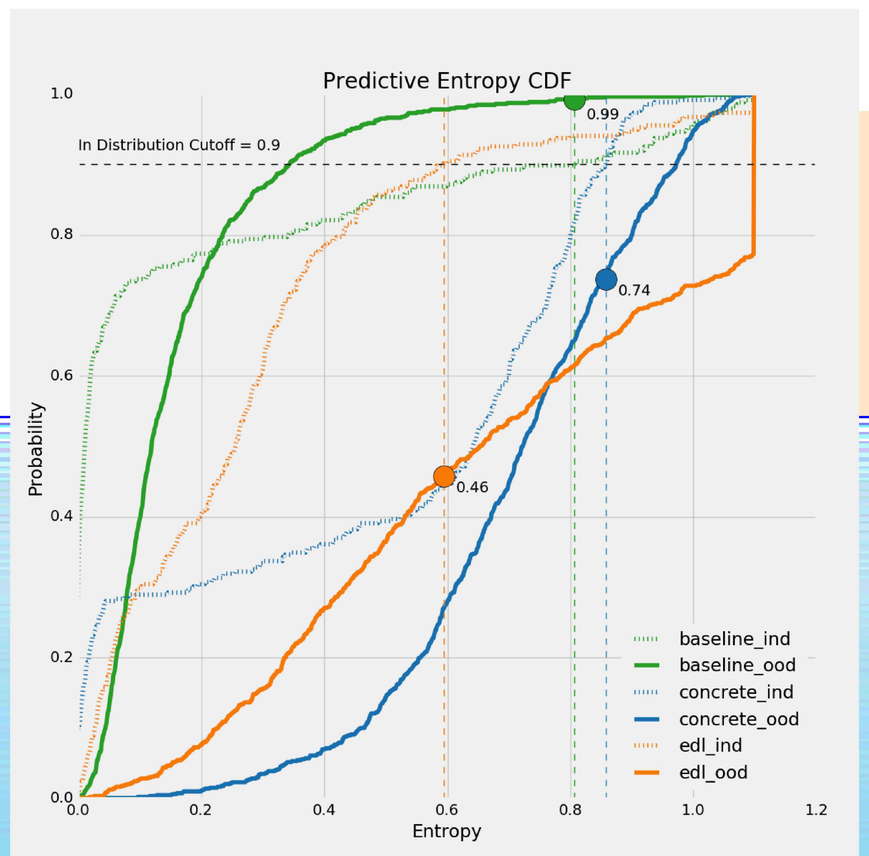


Evidentiary Deep Learning

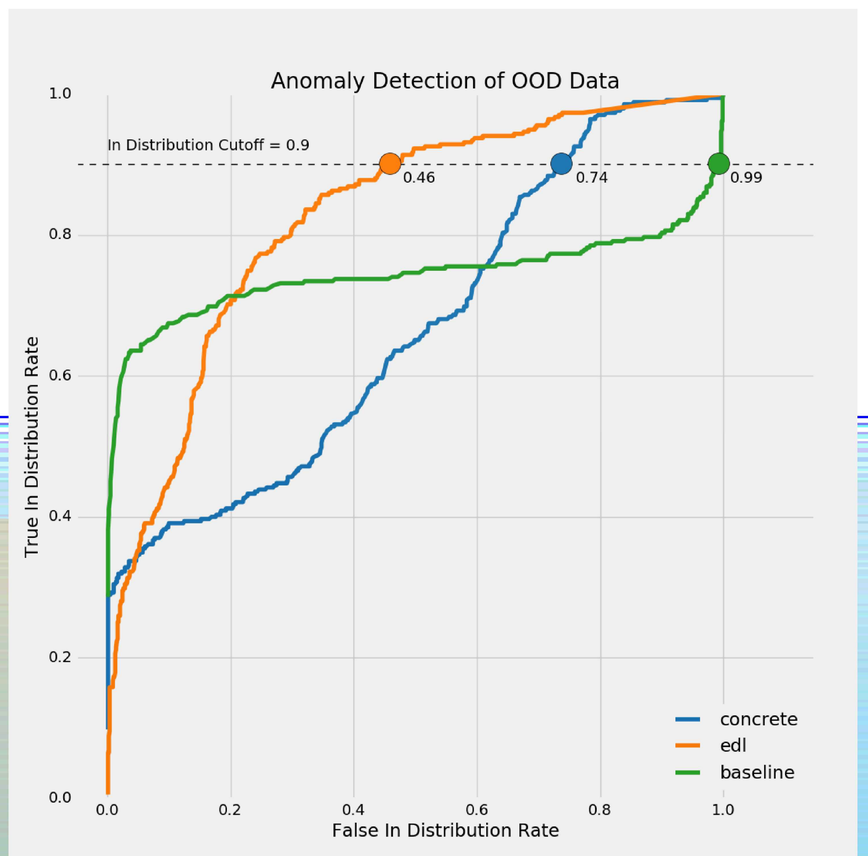


- Baseline softmax produces point estimate outputs. Concrete dropout and evidentiary deep learning methods are capable of producing uncertainty estimates.
- Baseline method produces much more confident output for all classes than either concrete dropout or evidentiary deep learning.

Results



- Predictive entropy is a good measure of how well out of distribution data is detected. Higher entropy (more uncertainty) is better.
- The farther apart the in and out of distribution data entropies are, the more effective an entropy cutoff will be in detecting anomalies.



U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. _____

Sandia National Laboratories