Advanced Data Mining Reading Course

Final Report

By

Vangjush Komini

vangjush@kth.se

June 8, 2021

1 Motivation

Deep learning (DL) has revolutionized the field of machine learning. This success is mainly attributed to the ability of this convolutional neural network (CNN) at generalizing over a larger amount of data than traditional methods without overfitting in most of the domain. Despite this recent success, DL methods are pretty incapable of representing uncertainty. Uncertainty is quite an essential ingredient for making DL more transparent as well as more robust. Softmax is not a sufficient mathematical mechanism to accommodate the inherent uncertainty, and the predecessor layers are all deterministic. Eventually, softmax alone will be capable of containing neither the models' nor the data uncertainty. Model (aleatoric) uncertainty is the subjective evaluation of the domain since one cannot sufficiently understand the data generation process. Data (epistemic) uncertainty is the noise presented in the data, which is not reducible even if you increase the data's volume. The first uncertainty is reducible, whereas the latter can is explainable by increasing the amount of data. One can legitly question the importance of uncertainty estimation (UE) in DL. UE offers the possibility to detect a distributional shift of both training data and labels. Out of distribution detection (OODD) is a scenario where the vanilla DL would systematically fail to accomplish. UE can offer the possibility of associating any test item with a likelihood metric which can then tell whether an item is OOD or not. Furthermore, UE can be beneficial at calibrating the predictive response of each model. Without UE, DL models cannot reflect the actual likelihood of the data distribution, resulting in a hectic behavior of the predictive response.

Most of the methods require multiple different runs of the machine learning model whereby aggregating the predictive outputs would eventually be possible to model the uncertainty. However, this is very computationally expensive, and getting the uncertainty within a single run is an active research direction one can undertake.

References