Master Thesis Defense Iswariya Manivannan

Date: 16.02.2021
Time: 10:00 AM
Room: Online

Title: Improving Uncertainty Estimates in Deep Learning Models using Knowledge Distillation

Abstract:
Ensembles of models are the state-of-the-art uncertainty estimation technique, yielding robust uncertainty measures and superior model performance. However, they have limited applications on resource constrained target platforms, as ensembling is a costly operation in terms of computational time and memory. These model deployment requirements are satisfied by using an approach called knowledge distillation, to transfer the knowledge from a cumbersome model or an ensemble of models to a smaller model. Recent approaches have used knowledge distillation to distill in the predictive distribution of ensemble outputs into a single model, consequently enabling it to produce the same, and sidestepping the need for using ensembles or other sampling based uncertainty estimation methods. Particularly, we experiment with two knowledge distillation based uncertainty estimation methods - hydra and prior networks, and show that the diversity of ensembles is not fully captured by these distilled models. Four new variants of the hydra model are introduced with the aim of improving its performance in terms of uncertainty quality, model accuracy, calibration and Out-Of-Distribution (OOD) detection. A detailed comparative analysis shows that the distilled hydra variant trained by using its hard predictions and soft teacher ensemble predictions, provides good uncertainty estimates next to that of ensembles. This hydra model produces high entropy predictions which lack in mutual information. A study of the effects of various distillation hyper-parameters reveals that the hydra models severely lack in diversity as compared to ensembles, and the temperature at which maximal performance is achieved highly depends on the noise in the dataset. The insights of this work also serve as a foundation for developing ensemble distillation methods that explicitly focus on retaining the diversity of ensembles in the distilled model. Additionally, a python based library consisting of different ensemble distillation based uncertainty estimation methods is developed.