Knowledge Distillation for Deep Neural Networks

Supervisor: Aleksei Triastcyn

**Keywords:** machine learning, deep learning, neural networks, knowledge distillation, model distillation

**Description**

When deploying machine learning models on mobile devices, such as phones and smart watches, a number of questions arises. They include computational efficiency and privacy issues. One of the ways to overcome these problems suggested in the literature is called knowledge distillation [1]. The method can be used to transfer the knowledge from a large model (a teacher) trained on a server into a smaller, compressed model (a student) simple enough to run on a device. The same approach has also been used to limit potential privacy loss due to exposure of the student model to an adversary [2].

In this project, you will learn about knowledge distillation and how to apply it to deep neural networks. You will implement the technique using a modern deep learning framework (e.g. tensorflow, mxnet, etc.) and apply it to MNIST [3] and SVHN [4] datasets (other datasets are also a possibility). Finally, you will analyse the trade-off between the amount of data used for training, accuracy, and differential privacy bounds.

**Project Objectives**

- Read and understand the papers on knowledge distillation.
- Implement a distillation algorithm and apply it to chosen datasets.
- Analyse the algorithm performance in terms of speed, accuracy, and privacy bounds.

**References**


