Optimization for Robust Deep Learning

M. Pawan Kumar, University of Oxford

State of the art neural networks have been shown to be highly susceptible to error under small deformations (the so-called adversarial examples). This severely limits their applicability to safety critical domains such as autonomous navigation and personalised medicine. To alleviate this deficiency, researchers have started to explore the idea of robust deep learning: iteratively finding deformations of training samples that cause an error, augmenting the training data set with the deformed samples, and retraining the network. To operationalize robust deep learning, we need to address two challenging optimization problems: (i) how do we find the error causing deformations; and (ii) how do we efficiently train the network. In this talk, I will give a brief overview of our novel proximal minimization based algorithms for the two aforementioned problems.