

**CRUNCH Seminars at Brown, Division of Applied Mathematics**

**Friday - December 4, 2020**

**Generalization effects of linear  
transformation in data augmentation**

**Sen Wu, Computer Science  
Department, Stanford University**

**Data augmentation is the technique of enlarging the training dataset through pre-defined transformation functions. By searching over a (possibly large) space of transformations through reinforcement learning based techniques, random sampling, and Bayesian optimization, data augmentation schemes have shown remarkable gains over various models on image and text classification tasks. Despite the rapid progress of these transformation search methods, precisely understanding their benefits remains a mystery due to the lack of analytic tools.**

**In this talk, I will introduce our study on when and why applying a family of linear transformations helps. Moreover, I will share an uncertainty-based random sampling scheme which, among the transformed data points, picks those ``providing the most information`` to improve the efficiency of transformation search procedures.**