Neural Networks for Cancer Survival Analysis Using High-Dimensional Data

Safoora Yousefi
Emory University

Abstract: Since the emergence of high throughput experiments such as Next Generation Sequencing, the volume of genomic data produced has been increasing exponentially. This data holds the key to accurate predictions of clinical outcomes and mapping patients to the optimal treatment. However, analyzing genomic data is challenged by its high-dimensionality. Many prediction methods face limitations in learning from high-dimensional data generated by these platforms, and rely on experts to hand-select a small number of features for training prediction models. In this thesis, we demonstrate how the latest advances in neural networks methods that have been remarkably successful in general high-dimensional prediction tasks can be leveraged to the problem of predicting cancer outcomes. We perform an extensive comparison of deep survival models and other state of the art machine learning methods for survival analysis. We appreciate that interpretability is of great importance in adapting neural networks in bioinformatics, and propose a framework for interpreting deep survival models using a risk back-propagation technique that can lead to new understanding of diseases. Finally, we illustrate that deep survival models can successfully transfer information across heterogeneous data sources to improve prognostic accuracy, and describe an adversarial multi-task learning approach that outperforms traditional multi-task learning methods. We provide an open-source software implementation of these frameworks that enables automatic training, evaluation and interpretation of deep survival models.

Wednesday, July 3, 2019, 2:30 pm
Mathematics and Science Center: MSC W301