PROJECT ABSTRACT

LEK-HENG LIM

The goal of this project is to build a complete picture of deep neural networks by piecing together current insights from algebraic, geometric, and topological studies. More specifically, it is to unify the geometric and toplogical views of a deep neural network as (i) a tropical rational map with decision boundary determined by its number of monomials, and (ii) a topology changing map that reduces the Betti numbers of a data set across its layers; and incorporate recently discovered algebraic properties like (a) the preservation of its expressive power when its weights are restricted to triangular, Toeplitz, or Hankel matrices, (b) the ill-posedness of the best k-layer approximation problem, and (c) equivariance under various group actions.

The main thrust of our technical approach comprises three interrelated prongs: (1) extend our preliminary study on classifiers to more complex AI models like generative adversarial networks, transformers, and variational autoencoders; (2) assimilate insight from the topological, geometric, and algebraic approaches to uncover the fundamental principles driving deep neural networks; (3) borrow new insights from neuroscience.

If successful, this will provide a theory of deep neural networks that could explain away the existing mysteries, allow for the design of more effective neural network architectures, and potentially unlock the door to explainable artificial intelligence and artificial general intelligence. More importantly, such a level of understanding will alleviate current fears that AI is an unpredictable and uncontrollable technology.

Military technologies are increasingly autonomated with AI and operate with minimal or no human intervention. However, ignorance of the underlying principles of deep neural networks impedes our ability to predict and control the behavior of such autonomous military technologies. Autonomous weapon systems, if out of control or if control is captured by an adversary, could cause catastrophic damage. Our inability to predict and control in turn impedes further automation of military technologies, due to the impending fear of building autonomous systems that could destroy humanity. Therein lies the defense relevance of this project: A genuine understanding of deep neural networks is indispensable for safe progress in military automation.