Ideally, when a neural network makes a wrong decision or encounters an out-of-distribution example, its predictive confidence should be as low as possible. Three primary contributions in this dissertation address this challenge. The first two contributions are new approaches to mitigate overconfident predictions in modern neural networks. In the first (1), called competitive overcomplete output layer neural networks, several classifiers, as part of the same output layer, are trained simultaneously and later their consensus produces more reliable predictions. The second approach (2) reformulates the original classification problem into several new versions by combining classes together and training a classifier on each. Experiments show that the resulting classifier aggregate, called fitted ensemble, is able to rectify predictive confidence values significantly better than conventional ensembles without sacrificing classification performance. Finally (3), a framework for evaluating the consistency of predictions called separable concept learning (SCL) is introduced. Together these contributions take a step towards achieving more reliable decisions under suboptimal conditions.

Major: Computer Science

Educational Career:
Bachelor's of Computer Engineering, BS, 2007, Ferdowsi University of Mashhad

Committee in Charge:
Kenneth O. Stanley, Chair, Computer Science
Ladislau Boloni, Computer Science
Fei Liu, Computer Science
Corey Bohil, Psychology

Approved for distribution by Kenneth O. Stanley, Committee Chair, on June 5, 2019.

The public is welcome to attend.