ReLU neuron hálózatok robusztusságának hatékony vizsgálata

Szász Attila
I. évf. programtervező informatikus mesterszak

Témavezető: Dr. Bánhelyi Balázs
SZTE TTIK Számítógépes Optimalizálás Tanszék

A mai neurális hálózatok nagy hatékonysággal képesek különböző osztályozási feladatok megoldására. A hálózatok pontossága az évek során folyamatosan nőtt, mely többek között az új tanítási technikák megjelenésének köszönhető. Számos kutatás kimutatta, hogy ezek a megbízhatónak vélt hálózatok is hibásan osztályozhatnak, gyakran már a tanító példákon megjelenő kis változtatások mellett. A szakirodalomban ezeket a mintákat adverzális/ellenséges példáknak nevezik. Az ilyen példák létezésének bizonyitására az évek során számos módszert dolgoztak ki, melyek egy része a létezésük megbízható kizárására, míg más részük az adverzális példák gyors megtalálására koncentrál. Munkánk során egy megbízhatóságot szem előtt tartó rendszert hoztunk létre, mely kezelhető futásidő növekedés mellett, hatékonyan végzi az egyes hálózatok adverzális mentességének bizonyítását.