

ReLU neuron hálózatok robusztusságának vizsgálata szimbolikus levezetéssel

Szász Attila

III. évf. programtervező informatikus

Témavezető: Dr. Bánhelyi Balázs

SZTE TTIK Számítógépes Optimalizálás Tanszék

A mai neurális hálók nagy hatékonysággal képesek különböző osztályozási feladatokat megoldani. Sajnos ezen osztályozás eredményére a tanítás során nem sok ráhatása van a felhasználónak. A jelenlegi vizsgálatokban észrevették, hogy előfordulnak olyan esetek, melyekben a neuronháló bár jól tanult, a tanító esetekhez hasonló, kis eltérést mutató példákon már hibázik. Ezeket a példákat nevezi a szakirodalom adverzális/ellenséges példáknak. Számos korszerű rendszert dolgoztak ki az évek során a kutatók, melyekkel az ilyen példák nem létezését lehetett bizonyítani. Egyes rendszerek a gyorsaságra, mások a megbízható számításokra koncentráltak. A munkánk során készítettünk egy, a megbízhatóságot szem előtt tartó kiértékelő rendszert, mely kimeneteit megvizsgáltuk különböző hálózatok esetében, illetve összehasonlítottuk más kiértékelési módszerekkel kapott eredményekkel.