

**Bendicsek Márton Bendegúz, BME VIK Mérnök informatika BSc, bendicsekb@gmail.com**  
**Konzulens: Dr. Renczes Balázs, adjunktus, Méréstechnika és Információs Rendszerek**  
**Tanszék (MIT)**

## **SZÁMÍTÓGÉPES KIÉRTÉKELÉSHEZ KAPCSOLÓDÓ NUMERIKUS PROBLÉMÁK**

### **ÖSSZEFOGLALÁS**

A cikk a neurális hálózatokban használt lebegőpontos számok kvantálási hibájából eredő számítási pontatlanságok analizálásával foglalkozik. A lebegőpontos számábrázolást az IEEE 754 szabvány szabályozza [2]. A cikkben bemutatom a kvantálási hiba modelljét B. Widrow és I. Kollár könyvéből [1]. A hibaterjedés illusztrálására megvizsgálom az összegzés során fellépő hibát, majd példákat hozok annak csökkentésére. A kutatás fő eredményeként bemutatom a perceptron hálózat kvantált aktivációkkal történő működését, melyhez megvizsgálom a Rosenblatt perceptront, mely képes lineárisan szeparálható problémák megtanulására. A cikkben ismertetem a hibaszámítás alapvető analizisének elvét és leírom a vizsgálatom eredményét.

### **ABSTRACT**

Diese Arbeit befasst sich mit der Analyse von Rechenungenauigkeiten, verursacht von Quantisierungsfehlern von Gleitkommazahlen, die in neuronalen Netzen verwendet werden. Die Darstellung der Gleitkommazahl ist durch den IEEE 754-Standard geregelt [2]. In dieser Arbeit stelle ich ein Modell des Quantisierungsfehlers aus dem Buch von B. Widrow und I. Kollár [1] vor. Um die Fehlerverteilung zu veranschaulichen, untersuche ich den Fehler, der während der Summierung auftritt, und gebe dann Beispiele, um ihn zu reduzieren. Als Hauptergebnis der Forschung präsentiere ich die Funktionsweise des Perzeptron-Netzwerks mit quantisierten Aktivierungen, für die ich das Perzeptron untersuche, das linear trennbare Probleme lernen kann. In dieser Arbeit beschreibe ich das Prinzip der Analyse der Fehlerberechnung und beschreibe das Ergebnis meiner Forschung.

### **BEVEZETÉS**

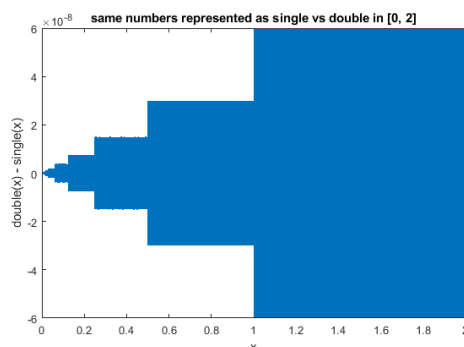
A neurális hálókat széles spektrumon alkalmazzák többek között regresszió, klasszifikáció, idősor előrejelzés, klaszteranalízis megvalósítására. Igen kurrens kutatási terület és számítási teljesítmény fejlődésével egyre nagyobb az alkalmazási területe is. A hagyományos jelfeldolgozással ötvözve viszont egy hasznos határterület, melynek segítségével például csökkenthető a hálók válaszája. A beszámolóban módszereket ismertetek a perceptron hálózat hibaszámításának meghatározásához.

### **KVANTÁLÁSI HIBA MODELLEZÉSE**

Az IEEE 754 szabvány egy kerekítési irány tulajdonságot rendel egy számábrázolás specifikációjához [2]. Egy számítás során a tagokat végtelenül pontosnak kell tekinteni, a számítás elvégzése után az eredményt (mely végtelenül pontos eddig a pontig) a szám

adott számrendszerbeli kerekítési irány tulajdonságának megfelelően kell kerekíteni. Ebből látható, hogy a számítások során például minden művelet után potenciálisan kerekítési hiba keletkezik.

Kerekítési hiba keletkezik akkor is, amikor egy számot lebegőpontos formátumban ábrázolunk. A kerekítési hiba (mely ez esetben a számaábrázolási hiba) legegyszerűbben az ábrázolt szám két különböző pontosságú leírásának különbségeként figyelhető meg. A nagyobb pontosságút (double [2]) tekintjük végtelenül pontosnak. Ezzel a módszerrel látható, hogy a kisebb pontosságú (single [2]) ábrázolás abszolútértékének növekedésével egyre nagyobb hiba keletkezik.



1. ábra IEEE double és single ábrázolás különbsége [2]

## ÖSSZEGZÉS SORÁN FELLÉPŐ KVANTÁLÁSI HIBA

Két lebegőpontos szám összeadásakor sok esetben keletkezik kerekítési hiba. Ennek elkerülésére két tag esetén az egyetlen lehetőség, hogy pontosabb számaábrázolást alkalmazunk. Ez a fejezet példákat mutat olyan módszerekre, amivel egy vektor összegzésénél a mantissza növelése nélkül is kisebb abszolút hibát eredményez.

### Összegzési módszerek a hiba csökkentésére

**Naiv összegzés:** az összegzendő elemeket sorban összeadjuk

**Páros összegzés:** az összegzendő elemeket párosával összeadjuk, majd rekurzívan párosával az eredményeket is összeadva jutunk az összeghez

**Kahan-féle kompenzált összegzés:** az összegzést úgy végezzük el, hogy minden lépésnél számontartjuk az adott összegzési lépés hibáját, mellyel a következő lépésnél kompenzálunk. A Kahan féle kompenzált összegzés általánosan pontosabbnak tekinthető a másik kettő módszernél, de a páros összeadás is látványos javulást hoz a naiv összegzéshez képest [3].

## **NEURÁLIS HÁLÓZATOK**

A mesterséges neurális hálózatok párhuzamos végrehajtásra képesek, fontos szempont a számítások sebessége. Képesek valamint tanulásra, ez a tanulási folyamat sokszor sok nagyságrenddel több időt vesz igénybe, mint egy eredmény kiszámítása.

Beágyazott rendszereknél és időkritikus alkalmazásokban a tanulási idő mellett azonban fontos lehet a hálózat válaszidejének lerövidítése is. Ez motiválta a következő fejezetek elemzéseit.

### **Klasszifikáció**

A klasszifikáció ellenőrzött tanulási módszer, melynek során a cél egy modell felépítése tanítópontok alapján, melyekre ismert a bemenet és az arra elvárt kimenet. Ezt a modellt később új adatpontok osztályozására lehet felhasználni. Klasszikus példa erre az időjárás-előrejelzés, vagy a hitelképesség eldöntése.

### **Neurális hálózat kvantált aktivációkkal**

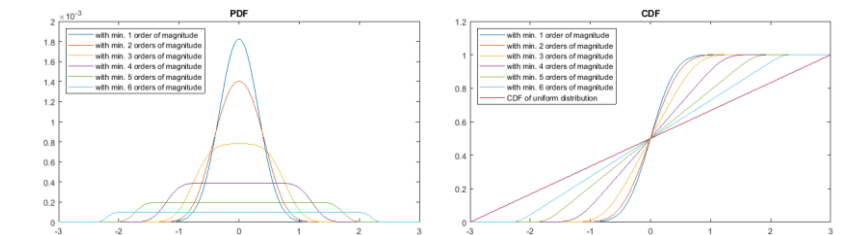
A következő fejezetben részletezett hálózat tanítása IEEE double pontossággal történt. Ebből a már betanított hálózatból kinyert súly- és eltolásvektorokat változatlan formában használtam fel. Tekintsük a double pontosságot végtelenül pontosnak, így ez a módszer azzal a feltételezéssel él, hogy a bemeneti értékek és a háló paraméterei végtelen pontossággal állnak rendelkezésünkre.

## **ROSENBLATT PERCEPTRON**

A Rosenblatt perceptron egy egy neuronból álló hálózat. Bemenetei az adatpont attribútumvektora, kimenete 1 vagy -1. A perceptron hálózat képes megtanulni egy lineárisan szeparálható problémát, azaz képes meghatározni azt a hipersíkot, amely az adatpontokat tartalmazó teret két részre osztja aszerint, hogy a pontok melyik osztályba tartoznak [4]. Látható, hogy ez egy klasszifikációs probléma, melyben két osztályt különböztetünk meg.

### **Perceptron hálózat kvantált aktivációkkal**

Az analízist kezdjük azzal, hogy az egyes műveletek utáni hiba varianciáját összegezzük. Ez az összegzési módszerek fejezetben tárgyaltak alapján történhet, hiszen a szorzásnak és az összeadásnak kumulálódik a varianciája. Egy kicsit kifinomultabb megoldás, hogy a hiba varianciájának 95-99 százalékát jelezzük, ez az egyenletes eloszlású valószínűségi változók összegéből számítható.



**2. ábra Egyenletes eloszlások egyesített, normált eloszlás és sűrűségfüggvényének alakulása**  
**Egynek nagyságrendekkel nagyobb a varianciája a többinél**

A 2. ábra segítségével könnyen látható, hogy egyenletes eloszlások összegzése során egy minimum 4 nagyságrenddel nagyobb variációjú már erősen elnyomja a többit. A közös sűrűség- és eloszlásfüggvény egy közel egyenletes eloszlás képét adja. Ez különösen igaz kevés számú valószínűségi változó összegére, ezért a 95 százalékos hibahatár nem ad sokkal jobb eredményt, mint a 99%, de mindkettő pontosabb, mint az egyszerűbb, variációk összegén alapuló előrejelzés. A kutatás során megvizsgáltam a hiba közelítését és a módszerek számításigényét is.

## KÖSZÖNETNYILVÁNÍTÁS

A cikkben ismertetett munka az Emberi Erőforrások Minisztérium EFOP-3.6.1-16-2016-00014 „Disruptív technológiák kutatás-fejlesztése az e-mobility területén és integrálásuk a mérnökképzésbe” c. pályázat keretében elnyert ösztöndíj támogatásával valósult meg.

## IRODALOM

1. B. Widrow, I. Kollár, "Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications", Cambridge University Press, Cambridge, UK, 2008. DOI: 10.1017/CBO9780511754661
2. IEEE Standard for Floating-Point Arithmetic," in IEEE Std 754-2019 (Revision of IEEE 754-2008), 2019 DOI: 10.1109/IEEESTD.2019.8766229
3. Nicholas J. Higham, "The Accuracy of Floating Point Summation", SIAM Journal on Scientific Computing 1993 14:4, 783-799, DOI: 10.1137/0914050
4. Altrichter Márta, Horváth Gábor, Pataki Béla, Strausz György, Takács Gábor, Valyon József, „Neurális hálózatok”, Panem, 2006, Magyarország ISBN-10 9-635454-64-3

## HIVATKOZÁSOK

BME: <http://www.bme.hu>