

FAOLLASHTIRISH FUNKSIYALARI: SIGMOID, TANH, RELU VA BOSHQALAR

Tojimamatov Israil Nurmamatovich

*Farg'ona davlat universiteti Amaliy matematika
va informatika kafedrası katta o'qituvchisi*

E-mail: israiltojimatov@gmail.com

Botiriva Maftuna Alisher qizi

*Farg'ona davlat universiteti Amaliy matematika
yo'nalishi 3-bosqich 23.07-guruh talabasi*

E-mail: xmmf1183@gmail.com

ANNOTATSIYA: Ushbu maqola sun'iy neyron tarmoq arxitekturalarida faollashtirish funksiyalarining har tomonlama tadqiqotiga va ularning chuqur o'rganish jarayonlaridagi muhim roliga bag'ishlangan. Tadqiqot klassik faollashtirish funksiyalarining matematik xususiyatlarini batafsil tahlil qilishni, jumladan sigmoid funksiyasi, giperbolik tangens, to'g'rilovchi chiziqli birlik va ushbu funksiyalarning zamonaviy modifikatsiyalarini o'z ichiga oladi. Tadqiqot neyron tarmoqlardagi noxiziqli o'zgarishlarning nazariy asoslarini, gradient tarqalish mexanizmlarini, yo'qolib borayotgan va portlovchi gradient muammolarini, shuningdek turli faollashtirish funksiyalarining hisoblash samaradorligini qamrab oladi.

Kalit so'zlar: faollashtirish funksiyalari, neyron tarmoqlar, chuqur o'rganish, sigmoid funksiyasi, giperbolik tangens, to'g'rilovchi chiziqli birlik, gradient tushish, xatoni orqaga tarqatish, noxiziqli o'zgarishlar, mashinali o'rganish

АННОТАЦИЯ: Настоящая статья посвящена всестороннему исследованию функций активации в архитектуре искусственных нейронных сетей и их критической роли в процессе глубокого обучения. В работе проводится детальный анализ математических свойств классических функций активации, включая сигмоидальную функцию, гиперболический тангенс, выпрямленную линейную единицу и современные модификации данных функций. Исследование охватывает теоретические основы нелинейных преобразований в нейронных сетях, механизмы градиентного распространения, проблемы исчезающего и взрывного градиента, а также вычислительную эффективность различных функций активации.

Ключевые слова: функции активации, нейронные сети, глубокое обучение, сигмоидальная функция, гиперболический тангенс, выпрямленная линейная единица, градиентный спуск, обратное распространение ошибки, нелинейные преобразования, машинное обучение

ANNOTATION: This article presents a comprehensive investigation of activation functions in artificial neural network architectures and their critical role in deep learning processes. The research provides detailed analysis of mathematical properties of classical activation functions, including sigmoid function, hyperbolic tangent, rectified linear unit, and contemporary modifications of these functions. The study encompasses theoretical foundations of nonlinear transformations in neural networks, gradient propagation mechanisms, vanishing and exploding gradient problems, as well as computational efficiency of various activation functions.

Keywords: activation functions, neural networks, deep learning, sigmoid function, hyperbolic tangent, rectified linear unit, gradient descent, backpropagation, nonlinear transformations, machine learning

KIRISH

Sun'iy intellekt va mashinali o'rganish sohasining rivojlanishi zamonaviy texnologik taraqqiyotning muhim qismiga aylandi. Neyron tarmoqlar hozirgi kunda tasvirlarni aniqlash, tabiiy tilni qayta ishlash, nutqni tanish, avtonom transport vositalari va tibbiy diagnostika kabi ko'plab amaliy sohalarda keng qo'llanilmoqda. Ushbu muvaffaqiyatlarning asosida chuqur neyron tarmoqlarning murakkab arxitekturalari va ularning tuzilish komponentlari yotadi.

Faollashtirish funksiyalari neyron tarmoqlarning fundamental tuzilma bloklari hisoblanadi va tarmoqqa nochiziqli xususiyatlarni berish orqali murakkab muammolarni hal qilish imkoniyatini yaratadi. Agar faollashtirish funksiyalari mavjud bo'lmasa, neyron tarmoq qanchalik chuqur bo'lmasin, faqat chiziqli o'zgarishlar ketma-ketligini ifodalash qobiliyatiga ega bo'lar edi. Bu esa tarmoqning ifodalash qobiliyatini jiddiy cheklagan bo'lar edi, chunki har qanday chiziqli funksiyalar kompozitsiyasi yana chiziqli funksiya hisoblanadi.

Faollashtirish funksiyalarining tarixi biologik neyronlar tadqiqoti bilan chambarchas bog'liq. Biologik neyronlar elektr signallarini qabul qiladi va ma'lum bir chegara qiymatiga erishilganda faollashadi, bu jarayon "hammasi yoki hech narsa" printsipiga asoslanadi. Ushbu biologik mexanizmni taqlid qilish orqali sun'iy neyron tarmoqlarda sigmoid va giperbolic tangens kabi funksiyalar joriy etildi. Ammo zamonaviy tadqiqotlar ko'rsatadiki, biologik neyronlarni to'liq taqlid qilish har doim ham optimal emas va ko'pincha soddaroq, ammo samaraliroq matematik funksiyalar yaxshiroq natijalar beradi.

Oxirgi yigirma yil ichida chuqur o'rganish sohasida katta yutuqlarga erishildi va bu yutuqlarning katta qismi faollashtirish funksiyalarini to'g'ri tanlash va optimallashtirishga bog'liq. To'g'rilovchi chiziqli birlik funksiyasi yoki ReLU ning paydo bo'lishi chuqur neyron tarmoqlarni o'qitishda inqilob yaratdi. ReLU ning

soddaligi va hisoblash samaradorligi uni zamonaviy chuqur o'rganish arxitekturalarida eng mashhur tanlovga aylantirdi.

Biroq, ReLU ham o'z kamchiliklariga ega va tadqiqotchilar doimiy ravishda yangi va yaxshilangan faollashtirish funksiyalarini ishlab chiqmoqdalar. Leaky ReLU, Parametric ReLU, Exponential Linear Unit, Swish va Mish kabi funksiyalar ReLU ning kamchiliklarini bartaraf etish va tarmoq samaradorligini oshirish uchun taklif etilgan. Har bir funksiya o'z afzalliklari va kamchiliklariga ega va muayyan vazifa uchun eng mos funksiyani tanlash muhim ahamiyatga ega.

Faollashtirish funksiyalarini tanlash gradient tarqalish jarayoniga bevosita ta'sir qiladi. Noto'g'ri tanlangan faollashtirish funksiyasi gradient yo'qolishi yoki gradient portlashi muammolariga olib kelishi mumkin, bu esa chuqur tarmoqlarni o'qitishni qiyinlashtiradi yoki imkonsiz qiladi. Shu sababli, faollashtirish funksiyalarining matematik xususiyatlarini, ularning hosilalarini va gradient oqimiga ta'sirini tushunish neyron tarmoq arxitekturalarini loyihalashda muhim ko'nikmalardan biridir.

Sigmoid funksiyasi

Sigmoid funksiyasi neyron tarmoqlar tarixida eng qadimgi va eng ko'p o'rganiladigan faollashtirish funksiyalaridan biridir. Ushbu funksiya biologik neyronlarning faollanish xususiyatlarini taqlid qilish uchun ishlab chiqilgan va logistik funksiya nomi bilan ham tanilgan. Sigmoid funksiyasining matematik ifodalanishi quyidagi ko'rinishga ega: kirish qiymatini noldan birga qadar bo'lgan oraliqqa akslantiradi va natija har doim musbat bo'ladi.

Sigmoid funksiyasining asosiy matematik xususiyatlari uning silliqligida va uzluksiz hosilaga ega ekanligidadir. Funksiya har bir nuqtada differentsiallanuvchan bo'lib, bu gradient asosidagi o'qitish algoritmlarini qo'llash imkonini beradi. Sigmoid funksiyasining hosilasi funksiyaning o'zi orqali ifodalanadi va bu hisoblashlarni soddalashtiradi. Hosilaning maksimal qiymati to'rttdan birga teng bo'lib, bu kirish qiymati nolga teng bo'lganda sodir bo'ladi.

Giperbolik tangens funksiyasi

Giperbolik tangens yoki tanh funksiyasi sigmoid funksiyasining yaxshilangan versiyasi hisoblanadi. Bu funksiya kirish qiymatini minus birdan plyus birgacha bo'lgan oraliqqa akslantiradi va sigmoid funksiyasidan farqli o'laroq, nol atrofida markazlashgan. Tanh funksiyasi matematik jihatdan sigmoid funksiyasining kengaytirilgan va siljigan shakli sifatida ifodalanishi mumkin.

Tanh funksiyasining eng muhim afzalligi uning nol atrofida simmetrik va markazlashgan ekanligidir. Bu xususiyat gradient tushish algoritmini samaraliroq qiladi, chunki gradientlar musbat va manfiy ishoralarga ega bo'lishi mumkin. Markazlashgan chiqish qiymatlari keyingi qatlamlardagi hisoblashlarni soddalashtiradi va o'qitish jarayonini tezlashtiradi. Tanh funksiyasining hosilasi ham funksiyaning o'zi orqali ifodalanadigan sodda shaklga ega.

Tanh funksiyasining hosilasi sigmoid funksiyasining hosilasidan kattaroqdir va maksimal qiymati birga teng. Bu sigmoid bilan solishtirganda kuchli gradientlarni ta'minlaydi va o'qitish jarayonini tezlashtiradi. Biroq tanh funksiyasi ham gradient yo'qolishi muammosiga duchor bo'ladi, garchi bu muammo sigmoid funksiyasiga nisbatan kamroq jiddiydir.

Klassik funksiyalarning qiyosiy tahlili

Sigmoid va tanh funksiyalarining qiyosiy tahlili bir qator muhim farqlarni ko'rsatadi. Tanh funksiyasi sigmoid funksiyasiga nisbatan kuchli gradientlarga ega va bu uni ko'pchilik vazifalar uchun yaxshiroq tanlovga aylantiradi. Markazlashgan chiqish qiymatlari tufayli tanh funksiyasi tarmoq samaradorligini oshiradi va o'qitish jarayonini tezlashtiradi.

Ikki funksiya ham silliq va uzluksiz hosilaga ega bo'lib, bu gradient asosidagi optimallashtirish algoritmlarini qo'llash uchun muhimdir. Biroq ularning hisoblash murakkabligi eksponensial operatsiyalar tufayli yuqori bo'lib qoladi. Zamonaviy dasturiy ta'minot va apparat vositalari bu operatsiyalarni tezlashtiradi, ammo ular hali ham sodda algebraik operatsiyalardan sekinroqdir.

ReLU ning zamonaviy modifikatsiyalari

ReLU funksiyasining kamchiliklarini bartaraf etish uchun tadqiqotchilar ko'plab modifikatsiyalarni taklif qildilar. Leaky ReLU funksiyasi o'lik neyronlar muammosini hal qilish uchun ishlab chiqilgan. Bu funksiya manfiy qiymatlar uchun nolga o'rniga kichik koeffitsientni qo'llaydi. Odatda bu koeffitsient o'ndan bir yoki yuzdan bir qiymatiga teng bo'ladi. Leaky ReLU manfiy qiymatlar uchun ham nolga teng bo'lmagan gradient beradi va bu o'lik neyronlar muammosini oldini oladi.

Parametric ReLU yoki PReLU funksiyasi Leaky ReLU ning umumlashtirilgan versiyasidir. PReLU da manfiy qiymatlar uchun koeffitsient o'rganiladigan parametr hisoblanadi. Bu koeffitsient tarmoq og'irliklari bilan birgalikda o'qitiladi va har bir neyron yoki kanal uchun alohida bo'lishi mumkin. Tadqiqotlar shuni ko'rsatadiki, PReLU ko'plab vazifalar uchun oddiy ReLU dan yaxshiroq natijalar beradi va gradient yo'qolishi muammosini yanada samaraliroq hal qiladi.

Adaptive va parametrik funksiyalar

Zamonaviy tadqiqotlar ko'rsatadiki, faollashtirish funksiyalari ham o'rganiladigan parametrlarga ega bo'lishi mumkin. Adaptive Piecewise Linear yoki APL funksiyasi kirish oralig'ini bir necha segmentlarga bo'ladi va har bir segmentda alohida chiziqli funksiya qo'llaniladi. Ushbu segmentlarning parametrlari o'qitish jarayonida o'rganiladi. APL funksiyasi juda moslashuvchan bo'lib, turli xil shakllarga moslashishi mumkin.

Maxout neyronlari ham adaptive faollashtirish g'oyasiga asoslangan. Maxout neyronida bir necha chiziqli transformatsiya qo'llaniladi va ularning maksimumi tanlanadi. Bu yondashuv faollashtirish funksiyasini tarmoq o'rganishi mumkin bo'lgan

parametrlar to'plami sifatida ifodalashga imkon beradi. Maxout neyronlari universal approksimatsilovchilar hisoblanadi va istalgan qavariq funktsiyani ifodalash qobiliyatiga ega.

Learnable Activation Functions yoki LAF lar faollashtirish funktsiyasini to'liq o'rganiladigan komponent sifatida ko'rib chiqadi. LAF lar turli usullar bilan amalga oshirilishi mumkin, jumladan spline interpolatsiyasi, neyron tarmoqlar yoki boshqa parametrik funktsiyalar yordamida. Bu yondashuv maksimal moslashuvchanlikni beradi, ammo qo'shimcha parametrlar va hisoblash xarajatlarini talab qiladi.

Adaptive faollashtirish funktsiyalarining afzalligi ularning muayyan vazifa uchun optimal shaklni avtomatik ravishda topish qobiliyatidir. Biroq bu yondashuv ortiqcha o'rganish xavfini oshiradi va ko'proq ma'lumotlar talab qiladi. Bundan tashqari, qo'shimcha parametrlar tarmoq hajmini oshiradi va o'qitish jarayonini murakkablashtiradi. Amalda, ko'pchilik hollarda standart funktsiyalar yetarli natijalar beradi.

Faollashtirish funktsiyalarining samaradorlik tahlili

Turli faollashtirish funktsiyalarining samaradorligini baholash uchun keng qamrovli eksperimentlar o'tkazildi. Ushbu eksperimentlar turli ma'lumotlar to'plamlari va turli tarmoq arxitekturalarida amalga oshirildi. Natijalar shuni ko'rsatadiki, faollashtirish funktsiyasining tanlovi vazifa turiga, tarmoq chuqurligiga va ma'lumotlar xususiyatlariga bog'liq.

Tasvirlarni tasniflash vazifalari uchun o'tkazilgan eksperimentlarda ReLU va uning modifikatsiyalari eng yaxshi natijalarni ko'rsatdi. CIFAR-10 va ImageNet kabi standart ma'lumotlar to'plamlarida PReLU va Mish funktsiyalari biroz ustunlik qildi. Biroq bu ustunlik ko'pincha kichik bo'lib, amaliy ahamiyati cheklangan. Oddiy ReLU funktsiyasi ko'pchilik hollarda yetarli natijalar beradi va uning soddaligi uni afzalroq tanlovga aylantiradi.

Tabiiy til qayta ishlash vazifalari uchun natijalar boshqacha bo'ldi. Transformer arxitekturalarida GELU funktsiyasi eng yaxshi natijalarni ko'rsatdi. BERT va GPT modellarida GELU ishlatish aniq yaxshilanishlarga olib keldi. Swish funktsiyasi ham yaxshi natijalar berdi, ammo GELU dan biroz pastroq ko'rsatkichlarga ega bo'ldi. Rekurrent neyron tarmoqlarda klassik tanh funktsiyasi hali ham samarali bo'lib qolmoqda.

Chuqur tarmoqlar uchun faollashtirish funktsiyasining tanlovi yanada muhim ahamiyatga ega. Yuz qatlamdan ortiq chuqurlikka ega tarmoqlarda gradient yo'qolishi muammosi jiddiyoq namoyon bo'ladi. Bunday holatlarda ELU, SELU va PReLU

funksiyalari ReLU dan ustunroq natijalar beradi. Biroq residual bog'lanishlar va normallashuv qatlamlari ishlatilganda, oddiy ReLU ham juda chuqur tarmoqlarda samarali bo'lishi mumkin.

Hisoblash samaradorligi va o'qitish tezligi

Faollashtirish funksiyalarining hisoblash samaradorligi amaliy qo'llanishlarda muhim omil hisoblanadi. ReLU funksiyasi eng tez hisoblanadigan funksiya bo'lib, u faqat oddiy solishtirish va tanlash operatsiyalarini talab qiladi. Sigmoid va tanh funksiyalari eksponensial hisoblashlar tufayli ancha sekinroqdir. Zamonaviy apparat akseleratorlarida eksponensial operatsiyalar optimallashtirilgan bo'lsa ham, ular hali ham ReLU dan sekinroq bo'lib qolmoqda.

Leaky ReLU va PReLU funksiyalari ReLU ga juda yaqin hisoblash xarajatlariga ega. Ular faqat bir qo'shimcha ko'paytirish operatsiyasini talab qiladi va bu zamonaviy protsessorlarda juda tez bajariladi. ELU va SELU funksiyalari eksponensial hisoblashlar tufayli biroz sekinroqdir, ammo ularning samaradorligi ba'zi holatlarda bu xarajatni oqlamoqda. Swish va Mish funksiyalari murakkab matematik ifodalari tufayli eng sekin funksiyalar hisoblanadi.

O'qitish tezligi nafaqat faollashtirish funksiyasining hisoblash samaradorligiga, balki uning gradient tarqalish xususiyatlariga ham bog'liq. ReLU funksiyasi ko'pincha eng tez yaqinlashishni ta'minlaydi, ayniqsa maqbul o'rganish tezligi tanlanganida. PReLU va ELU funksiyalari ba'zi holatlarda yanada tezroq yaqinlashishni beradi. SELU funksiyasi o'z-o'zidan normallashuv xususiyati tufayli ba'zi arxitekturalarda o'qitish tezligini sezilarli darajada oshiradi.

Gradient yo'qolishi va gradient portlashi muammolari o'qitish tezligiga katta ta'sir ko'rsatadi. Klassik sigmoid va tanh funksiyalari juda chuqur tarmoqlarda o'qitishni sekinlashtiradi yoki to'xtatadi. ReLU va uning modifikatsiyalari bu muammoni bartaraf etadi va chuqur tarmoqlarni samarali o'qitish imkonini beradi. Gradient qirqish va normallashuv texnikalari bilan birgalikda zamonaviy faollashtirish funksiyalari juda chuqur tarmoqlarni muvaffaqiyatli o'qitish imkonini beradi.

Matn ma'lumotlari va tabiiy til qayta ishlash vazifalari uchun transformer arxitekturalari GELU funksiyasi bilan eng yaxshi natijalarni ko'rsatadi. Rekurrent neyron tarmoqlarda tanh funksiyasi hali ham afzalroq tanlov bo'lib qolmoqda, ayniqsa LSTM va GRU arxitekturalarida. Biroq ba'zi zamonaviy rekurrent arxitekturalarda ReLU ham sinab ko'rilmogda va qo'shimcha normallashuv bilan yaxshi natijalar beradi.

Vaqt qatorlarini bashorat qilish vazifalari uchun tanh va ReLU kombinatsiyasi ko'pincha yaxshi natijalar beradi. Vaqt qatorlarining xususiyatlari va vazifaning talablariga qarab turli faollashtirish funksiyalarini sinab ko'rish tavsiya etiladi. Attention mexanizmlari vaqt qatorlari uchun ham samarali bo'lib, bu holda GELU yoki Swish funksiyalari qo'llaniladi.

Audio signallarni qayta ishlashda konvolyutsion tarmoqlar keng qo'llaniladi va ReLU asosiy tanlov hisoblanadi. Spektrogrammalar va boshqa audio tasvirlari uchun tasvir tarmoqlari arxitekturalari moslashtiriladi va ular uchun ham ReLU yaxshi natijalar beradi. Generativ modellar uchun Leaky ReLU va ELU funktsiyalari ko'proq qo'llaniladi.

XULOSA

Faollashtirish funktsiyalari neyron tarmoqlarning muhim tuzilma bloklari hisoblanadi va ularning to'g'ri tanlovi tarmoq samaradorligiga katta ta'sir ko'rsatadi. Ushbu maqolada biz klassik va zamonaviy faollashtirish funktsiyalarini batafsil ko'rib chiqdik, ularning matematik xususiyatlarini tahlil qildik va amaliy qo'llanilishlarini muhokama qildik. Tadqiqot natijalari shuni ko'rsatadiki, faollashtirish funktsiyasining tanlovi vazifa turiga, ma'lumotlar xususiyatlariga va tarmoq arxitekturasiga bog'liq.

Sigmoid va tanh kabi klassik funktsiyalar neyron tarmoqlar tarixida muhim rol o'ynagan va hozirda ham ba'zi maxsus vazifalarda qo'llanilmoqda. Ushbu funktsiyalar silliq va interpretatsiya qilish uchun oson bo'lsa ham, gradient yo'qolishi muammosi va hisoblash samaradorligining pastligi ularning qo'llanilishini cheklaydi. Zamonaviy chuqur o'rganish arxitekturalarida bu funktsiyalar asosan chiqish qatlamlarida va rekurrent tarmoqlarda ishlatiladi.

ReLU funktsiyasining paydo bo'lishi chuqur o'rganish sohasida muhim yutuq bo'ldi. ReLU ning soddaligi, hisoblash samaradorligi va gradient yo'qolishi muammosini musbat qiymatlar uchun bartaraf etish qobiliyati uni eng mashhur faollashtirish funktsiyasiga aylantirdi. Biroq ReLU ning o'lik neyronlar muammosi uning kamchiligi hisoblanadi va bu muammo ko'plab modifikatsiyalarning paydo bo'lishiga sabab bo'ldi.

FOYDALANILGAN ADABIYOTLAR

1. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016.
2. Glorot X., Bengio Y. "Understanding the difficulty of training deep feedforward neural networks." AISTATS, 2010.
3. Ramachandran P., Zoph B., Le Q.V. "Swish: a Self-Gated Activation Function." Google Brain, 2017.
4. He K., Zhang X., Ren S., Sun J. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification." ICCV, 2015.
5. Clevert D., Unterthiner T., Hochreiter S. "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)." ICLR, 2016.
6. Hendrycks D., Gimpel K. "Gaussian Error Linear Units (GELUs)." arXiv:1606.08415, 2016.
7. Maas A.L., Hannun A.Y., Ng A.Y. "Rectifier Nonlinearities Improve Neural Network Acoustic Models." ICML, 2013.
8. Haykin S. *Neural Networks and Learning Machines*. Pearson, 2009.