



NEYRON TARMOQLARDA GRADIENT YO'QOLISHI  
MUAMMOSI

*Tursunaliyeva Mohinur Zoirjon qizi*

*Farg'ona davlat universiteti talabasi*

[ORCID ID: 0009-0006-3514-8787](https://orcid.org/0009-0006-3514-8787)

**Annotatsiya:** Maqolada chuqur neyron tarmoqlarda yuzaga keladigan gradient yo'qolishi muammosi tahlil qilinadi. Uning matematik mohiyati, sabab va tarmoq o'qitilishiga ta'siri ko'rib chiqilgan. Shuningdek, ReLU aktivatsiyasi, Batch Normalization, Residual tarmoqlar va optimal vazn boshlang'ich qiymatlash metodlari yordamida muammoni kamaytirish usullari taqdim etilgan. Tadqiqot natijalari chuqur o'rganish samaradorligini oshirishga yordam beradi.

**Аннотация:** В статье рассматривается проблема исчезающего градиента в глубоких нейронных сетях. Проанализированы её причины, математическая основа и влияние на обучение. Обсуждаются методы снижения проблемы: ReLU, Batch Normalization, Residual Networks и инициализация весов. Результаты показывают улучшение эффективности обучения.

**Abstract:** The article studies the vanishing gradient problem in deep neural networks. Its causes, mathematical basis, and impact on training are analyzed. Methods to mitigate the issue, including ReLU, Batch Normalization, Residual Networks, and weight initialization, are discussed. Results show improved training efficiency.

**Kalit so'zlar:** Chuqur o'rganish, Neyron tarmoqlar, Gradient yo'qolishi, Backpropagation, Aktivatsiya funksiyalari, ReLU, Batch Normalization, Residual tarmoqlar, Vazn inicializatsiyasi, Model samaradorligi, O'qitish jarayoni.

**Ключевые слова:** Глубокое обучение, Нейронные сети, Исчезающий градиент, Обратное распространение ошибки, Функции активации, ReLU,



*Batch Normalization, Резидентные сети, Инициализация весов, Эффективность модели, Процесс обучения.*

**Keywords:** *Deep learning, Neural networks, Vanishing gradient, Backpropagation, Activation functions, ReLU, Batch Normalization, Residual networks, Weight initialization, Model efficiency, Training process.*

## KIRISH

So‘nggi yillarda sun‘iy intellekt va chuqur o‘rganish texnologiyalari ilm-fan va texnologiyaning eng tez rivojlanayotgan sohalaridan biriga aylandi. Chuqur o‘rganish usullari ko‘pincha ko‘p qatlamli neyron tarmoqlardan foydalanishga asoslanadi. Bunday tarmoqlar tasvirni aniqlash, nutqni qayta ishlash, tabiiy tilni tushunish, tibbiy diagnostika va boshqa ko‘plab murakkab vazifalarni hal qilishda katta muvaffaqiyatlarga erishgan. Neyron tarmoqlarning asosiy afzalligi shundaki, ular katta hajmdagi ma‘lumotlardan murakkab bog‘lanishlarni avtomatik ravishda o‘rganish qobiliyatiga ega.

Biroq chuqur neyron tarmoqlarni samarali o‘qitish jarayoni ko‘plab matematik va hisoblash muammolari bilan bog‘liq. Eng muhim va keng muhokama qilinadigan muammolardan biri — **gradient yo‘qolishi muammosi** hisoblanadi. Bu muammo, asosan ko‘p qatlamli neyron tarmoqlarni o‘qitishda yuzaga keladi va modelning samarali o‘rganish jarayoniga jiddiy to‘sqinlik qiladi.

Neyron tarmoqlarni o‘qitish jarayoni odatda, xatolikni orqaga tarqatish algoritmi asosida amalga oshiriladi. Bu jarayonda modelning xatoligi chiqish qatlamidan boshlab, oldingi qatlamlarga tomon uzatiladi va har bir qatlamning vazn koeffitsiyentlari shu xatolik asosida yangilanadi. Bunda asosiy rol ni gradient — ya‘ni xatolik funksiyasining vaznlarga nisbatan hosilasi bajaradi. Gradient qiymatlari yordamida vaznlarning qaysi yo‘nalishda o‘zgartirilishi kerakligi aniqlanadi.

Ammo neyron tarmoq qatlamlari soni ortib borgani sari ayrim holatlarda gradient qiymatlari juda kichik bo‘lib boradi. Natijada, gradient signalining oldingi qatlamlarga yetib borishi sezilarli darajada kamayadi yoki deyarli yo‘qolib ketadi. Aynan shu holat **gradient yo‘qolishi muammosi** deb ataladi. Gradient qiymatlari juda kichik bo‘lganda, tarmoqning dastlabki qatlamlari deyarli o‘rganmaydi, ya‘ni



ularning vaznlari o'zgarmay qoladi. Bu esa, modelning umumiy o'qitilish samaradorligini pasaytiradi va chuqur tarmoqlarni o'rganishni qiyinlashtiradi.

Gradient yo'qolishi muammosi, ayniqsa sigmoidal yoki giperbolik tangens kabi ba'zi aktivatsiya funksiyalaridan foydalanilganda ko'proq uchraydi. Ushbu funksiyalar hosilasining kichik qiymatlari qatlamlar bo'ylab ko'payib borib, natijada gradient qiymatlari tezda nolga yaqinlashib ketishi mumkin. Shu sababli chuqur neyron tarmoqlarni samarali o'qitish uchun ushbu muammoni chuqur o'rganish va uni kamaytirishga qaratilgan usullarni ishlab chiqish muhim ilmiy vazifa hisoblanadi.

Hozirgi kunda gradient yo'qolishi muammosini kamaytirish uchun turli metodlar taklif qilingan. Jumladan, yangi aktivatsiya funksiyalaridan foydalanish, maxsus arxitekturalar yaratish, qatlamlarni normallashtirish usullari va optimallashtirish algoritmlarini takomillashtirish kabi yondashuvlar keng qo'llanilmoqda. Shu sababli, gradient yo'qolishi muammosini o'rganish chuqur o'rganish tizimlarining samaradorligini oshirishda muhim ilmiy va amaliy ahamiyatga ega.

## ASOSIY QISM

Neyron tarmoqlarni o'qitish jarayonida model parametrlarini yangilash xatolikni orqaga tarqatish algoritmi orqali amalga oshiriladi. Ushbu algoritmda tarmoqning chiqishida hisoblangan xatolik qiymati tarmoq qatlamlari bo'ylab orqaga uzatiladi va har bir qatlamning vazn koeffitsiyentlari mos ravishda yangilanadi.

Agar neyron tarmoqda  $L$  ta qatlam mavjud bo'lsa, unda  $\ell$ -qatlamdagi vaznlar  $W^{(\ell)}$  bilan belgilanadi. Xatolik funksiyasi  $J(W)$  bo'lsa, vaznlarni yangilash quyidagi gradient tushunchasiga asoslanadi:

$$\frac{\partial J}{\partial W^{(\ell)}}$$

Bu ifoda xatolik funksiyasining vaznlarga nisbatan hosilasini bildiradi. Gradient qiymati yordamida vaznlar quyidagicha yangilanadi:

$$W^{(\ell)} = W^{(\ell)} - \eta \frac{\partial J}{\partial W^{(\ell)}}$$



bu yerda:

- $\eta$  — o'rganish tezligi,
- $\frac{\partial J}{\partial W(\ell)}$  — gradient qiymati.

Chuqur neyron tarmoqlarda gradientni hisoblash jarayonida **zanjir qoidasidan** foydalaniladi. Agar  $\ell$  -qatlamdagi aktivatsiya funksiyasi  $f(\cdot)$  bo'lsa, gradient quyidagi ko'rinishda ifodalanadi:

$$\frac{\partial J}{\partial W(\ell)} = \frac{\partial J}{\partial a(L)} \prod_{k=\ell}^{L-1} f'(z^{(k)})$$

bu yerda:

- $f'(z^{(k)})$  — aktivatsiya funksiyasining hosilasi,
- $L$  — tarmoqning oxirgi qatlami.

Agar aktivatsiya funksiyasi hosilasi qiymatlari  $0 < f'(z) < 1$  oralig'ida bo'lsa (masalan sigmoid funksiyada), u holda qatlamlar bo'ylab bu qiymatlar ko'payib boradi. Natijada ko'p qatlamli tarmoqlarda ushbu ko'paytma juda kichik qiymatga ega bo'lib qoladi:

$$f'(z_1) \cdot f'(z_2) \cdot \dots \cdot f'(z_n) \approx 0$$

Natijada gradient qiymati nolga yaqinlashadi. Bu esa tarmoqning dastlabki qatlamlarida vaznlarning deyarli o'zgarmasligiga olib keladi. Shuning uchun modelning chuqur qatlamlari samarali o'rganmaydi. Aynan shu holat **gradient yo'qolishi muammosi** deb ataladi.

Gradient yo'qolishi muammosi bir necha asosiy omillar bilan bog'liq:

**1. Aktivatsiya funksiyalarining xususiyati.** Sigmoid yoki giperbolik tangens funksiyalarining hosilalari juda kichik qiymatlarga ega bo'lishi mumkin. Bu gradientning tez kamayishiga olib keladi.

**2. Tarmoq chuqurligining ortishi.** Qatlamlar soni ortgan sari gradient qiymatlari ko'paytmasi juda kichik sonlarga aylanishi mumkin.

**3. Vaznlarning noto'g'ri boshlang'ich qiymatlari.** Agar vaznlar juda kichik yoki juda katta qiymatlarda boshlansa, gradientlar tez so'nib qolishi mumkin.

Ushbu muammolar quyidagi salbiy oqibatlarga olib keladi:



- Dastlabki qatlamlar o'rganmay qoladi.
- Modelning o'qitish jarayoni juda sekinlashadi.
- Tarmoq murakkab bog'lanishlarni o'rganishda qiynaladi.
- Modelning aniqligi pasayadi.

Natijada chuqur neyron tarmoqlarning samaradorligi sezilarli darajada kamayadi.

Hozirgi zamonaviy chuqur o'rganish tizimlarida ushbu muammoni kamaytirish uchun bir nechta yondashuvlar qo'llaniladi: **1) ReLU aktivatsiya funksiyasi:**

Rectified Linear Unit funksiyasi gradientning yo'qolish ehtimolini kamaytiradi; **2)**

**Batch Normalization usuli:** Har bir qatlamdagi ma'lumotlarni normallashtirish orqali o'qitish jarayonini barqarorlashtiradi. **3) Residual tarmoqlar (ResNet):**

Maxsus skip-connection strukturalari gradientning chuqur qatlamlarga yetib borishini osonlashtiradi. **4) To'g'ri vazn boshlang'ich qiymatlari:** Xavier yoki He initialization usullari gradientning stabil bo'lishiga yordam beradi.

## MUHOKAMA

Chuqur neyron tarmoqlarni o'qitish jarayonida gradient yo'qolishi muammosi sun'iy intellekt tadqiqotlarida eng muhim nazariy va amaliy masalalardan biri hisoblanadi. Ushbu muammo asosan ko'p qatlamli neyron tarmoqlarda yuzaga kelib, model parametrlarini samarali yangilash jarayoniga salbiy ta'sir ko'rsatadi. Gradient qiymatlarining qatlamlar bo'ylab kamayib borishi natijasida tarmoqning dastlabki qatlamlari deyarli o'rganmay qoladi. Bu holat chuqur tarmoqlarning asosiy afzalligi bo'lgan murakkab xususiyatlarni aniqlash imkoniyatini cheklaydi.

Muhokamalar shuni ko'rsatadiki, gradient yo'qolishi muammosi asosan aktivatsiya funksiyalarining matematik xususiyatlari, tarmoq chuqurligining ortishi va parametrlarning boshlang'ich qiymatlariga bog'liq holda yuzaga keladi. Masalan, sigmoid va giperbolik tangens funksiyalarining hosilalari kichik qiymatlarga ega bo'lgani sababli, gradientlar qatlamlar bo'ylab tez kamayadi. Natijada o'qitish jarayoni samarasiz bo'lib qoladi.



Zamonaviy tadqiqotlar esa ushbu muammoni kamaytirish uchun turli yondashuvlarni taklif etmoqda. Jumladan, ReLU kabi yangi aktivatsiya funksiyalaridan foydalanish gradientning barqaror saqlanishiga yordam beradi. Shuningdek, Batch Normalization texnologiyasi tarmoq qatlamlaridagi ma'lumotlarni normallashtirish orqali o'qitish jarayonini tezlashtiradi. Bundan tashqari, Residual neyron tarmoqlari (ResNet) kabi yangi arxitekturalar gradientning chuqur qatlamlarga yetib borishini osonlashtirib, chuqur tarmoqlarni samarali o'qitishga imkon yaratadi.

Shu sababli gradient yo'qolishi muammosini o'rganish nafaqat nazariy jihatdan, balki amaliy jihatdan ham muhim hisoblanadi. Chunki ushbu muammoni hal qilish orqali yanada chuqur va murakkab neyron tarmoqlarni yaratish, ularning aniqligini oshirish hamda turli real masalalarni samarali yechish mumkin bo'ladi.

## NATIJALAR

Tadqiqotlar natijasida neyron tarmoqlarda gradient yo'qolishi muammosining yuzaga kelish sabablari va uning o'qitish jarayoniga ta'siri tahlil qilindi. Olingan natijalar shuni ko'rsatadiki, chuqur neyron tarmoqlarda gradient qiymatlarining juda kichik bo'lib qolishi tarmoqning dastlabki qatlamlarida o'rganish jarayonining sustlashishiga olib keladi. Bu esa modelning umumiy samaradorligini pasaytiradi.

Shuningdek, tahlillar natijasida gradient yo'qolishi muammosini kamaytirish uchun zamonaviy usullar, jumladan ReLU aktivatsiya funksiyasi, Batch Normalization, Residual arxitekturalar hamda to'g'ri vaznlarni boshlang'ich qiymatlash metodlarining samarali ekanligi aniqlangan. Ushbu usullar yordamida gradientlarning qatlamlar bo'ylab barqaror uzatilishi ta'minlanadi va chuqur neyron tarmoqlarni o'qitish samaradorligi sezilarli darajada oshadi.

Natijada gradient yo'qolishi muammosini chuqur o'rganish va uni bartaraf etishga qaratilgan usullarni qo'llash sun'iy intellekt tizimlarining aniqligi, barqarorligi va ishlash samaradorligini oshirishda muhim ahamiyat kasb etadi.

## XULOSA



Ushbu tadqiqotda neyron tarmoqlarda **gradient yo'qolishi muammosining** nazariy asoslari, uning yuzaga kelish sabablari hamda chuqur o'rganish jarayoniga ta'siri tahlil qilindi. Tahlillar shuni ko'rsatdiki, ko'p qatlamli neyron tarmoqlarda gradient qiymatlarining qatlamlar bo'ylab kamayib borishi model parametrlarining samarali yangilanmasligiga olib keladi. Natijada tarmoqning dastlabki qatlamlari yetarli darajada o'rganmaydi va modelning umumiy o'qitish samaradorligi pasayadi.

Shuningdek, tadqiqot jarayonida gradient yo'qolishi muammosi asosan aktivatsiya funksiyalarining matematik xususiyatlari, tarmoq chuqurligining ortishi hamda vaznlarning boshlang'ich qiymatlariga bog'liq holda yuzaga kelishi aniqlanadi. Ushbu muammoni kamaytirish uchun zamonaviy chuqur o'rganish tizimlarida ReLU aktivatsiya funksiyasi, Batch Normalization usuli, Residual arxitekturalar va optimal vazn boshlang'ich qiymatlash metodlari keng qo'llanilayotgani ko'rsatildi. Bu yondashuvlar gradientlarning qatlamlar bo'ylab barqaror uzatilishini ta'minlab, chuqur neyron tarmoqlarning o'qitish jarayonini sezilarli darajada yaxshilaydi.

Shu bilan birga, kelajakdagi tadqiqotlar uchun yangi yo'nalish sifatida adaptiv aktivatsiya funksiyalarini ishlab chiqish va gradient oqimini avtomatik nazorat qiluvchi neyron arxitekturalarni yaratish muhim ilmiy vazifa bo'lib qolmoqda. Bunday yondashuvlar neyron tarmoqlarning o'rganish jarayonini yanada barqarorlashtirishi va juda chuqur modellarni samarali o'qitish imkoniyatini kengaytirishi mumkin.

Natijada gradient yo'qolishi muammosini chuqur o'rganish va uni bartaraf etishga qaratilgan yangi metodlarni ishlab chiqish sun'iy intellekt tizimlarining aniqligi, barqarorligi hamda amaliy qo'llanilish samaradorligini oshirishda muhim ahamiyat kasb etadi.

## FOYDALANILGAN ADABIYOTLAR:

1. Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. MIT Press, 2016.
2. Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.



3. Simon Haykin. Neural Networks and Learning Machines. Pearson Education, 2009.
4. Michael Nielsen. Neural Networks and Deep Learning. Determination Press, 2015.
5. Xavier Glorot, Yoshua Bengio. Understanding the Difficulty of Training Deep Feedforward Neural Networks. Proceedings of AISTATS, 2010.
6. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. Proceedings of CVPR, 2015.
7. Sergey Ioffe, Christian Szegedy. Batch Normalization Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of ICML, 2015.
8. Jürgen Schmidhuber. Deep Learning in Neural Networks An Overview. Neural Networks Journal, 2015.