XATOLARNI TESKARI TARQATISH USULI(BACK-PROPAGATION)

Raximov Quvvatali

Farg'ona Davlat Universiteti amaliy matematika va informatika kafedrasi mudiri,texnika fanlari bo'yicha falsafa doktori(PhD).

quvvatali.rahimov@gmail.com

Azimjonova Mohinur Asiljon qizi

Farg'ona Davlat Universiteti 2-kurs talabasi,

azimjonovamohinur88@gmail.com

Annotatsiya: Neyron tarmoqlari gradient tushishining ba'zi modifikatsiyalari yordamida o'qitiladi va uni qo'llash uchun siz barcha o'quv parametrlari bo'yicha yo'qotish funktsiyasining gradientlarini samarali hisoblashingiz kerak. Ko'rinib turibdiki, ba'zi bir chalkash hisoblash grafigi uchun bu juda qiyin vazifa bo'lishi mumkin, ammo xatoni orqaga qaytarish usuli yordamga keladi.

Kalit so'zlar: Perseptron, neyron, gradient, model.

KIRISH

Neyron tarmoqlari gradient tushishining ba'zi modifikatsiyalari yordamida o'qitiladi va uni qo'llash uchun siz barcha o'quv parametrlari bo'yicha yo'qotish funktsiyasining gradientlarini samarali hisoblashingiz kerak. Ko'rinib turibdiki, ba'zi bir chalkash hisoblash grafigi uchun bu juda qiyin vazifa bo'lishi mumkin, ammo xatoni orqaga qaytarish usuli yordamga keladi.

Xatolarning teskari tarqalish usulini kashf etish sun'iy intellekt sohasidagi eng muhim voqealardan biriga aylandi. U 1986 yilda Devid E. Rumelxart, Jeffri E. Xinton va Ronald J. Uilyam va mustaqil ravishda va bir vaqtning o'zida Krasnoyarsk matematiklari S. I. Bartsev va V. A. Ohonin tomonidan. O'shandan beri neyron tarmoq parametrlari gradientlarini topish uchun murakkab funksiyaning hosilasini hisoblash usuli qo'llanilmoqda va tarmoq parametrlari gradientlarini baholash hech bo'lmaganda murakkab muhandislik vazifasiga aylandi, ammo endi san'at emas. Amaldagi matematik apparatning soddaligiga qaramay, ushbu usulning paydo bo'lishi sun'iy neyron tarmoqlarining rivojlanishida sezilarli sakrashga olib keldi.

Usulning mohiyatini murakkab funktsiyaning hosilasi formulasidan ahamiyatsiz bo'lgan bitta formula bilan yozish mumkin:

$$f(x) = g_m(g_{m-1}(\ldots(g_1(x))\ldots))$$

agar

$$\frac{\partial f}{\partial x} = \frac{\partial g_m}{\partial g_{m-1}} \frac{\partial g_{m-1}}{\partial g_{m-2}} \ldots \frac{\partial g_2}{\partial g_1} \frac{\partial g_1}{\partial x}$$

$$f'(w_0) = g'_m(g_{m-1}(\ldots g_1(w_0)\ldots)) \;\; g'_{m-1}(g_{m-2}(\ldots g_1(w_0)\ldots))$$

$f(x) = g_m(g_{m-1}(\ldots(g_1(x))\ldots))$ keyin Endi biz gradientlarni ketma - ket, bitta orqaga o'tish paytida hisoblash mumkinligini ko'rmoqdamiz $\frac{\partial g_m}{\partial g_{m-1}}$ va har safar oldingi qatlamning qisman hosilalariga ko'paytirish.

Bir o'lchovli holatda, narsalar ayniqsa sodda ko'rinadi. Ruxsat bering w_0 - biz farqlamoqchi bo'lgan o'zgaruvchi va murakkab funktsiya shaklga ega

hamma qaerda g_i skalyar. Keyin

$$f'(w_0) = g'_m(g_{m-1}(...g_1(w_0)...))\ \ g'_{m-1}(g_{m-2}(...g_1(w_0)...))$$

$$f(w_0) = g_m(g_{m-1}(...g_1(w_0)...))$$

$$g_i$$

$$g_1(w_0), g_2(g_1(w_0)), ..., g_{m-1}(...g_1(w_0)...)$$

$$w_i$$

$$\frac{\partial f}{\partial w_0} = (-y)\ \ e^{-y(w_0+w_1x_1+w_2x_2)}\ \ \frac{-1}{1+e^{-y(w_0+w_1x_1+w_2x_2)}}$$

Ushbu formulaning mohiyati quyidagicha. Agar biz allaqachon oldinga o'tishni amalga oshirgan bo'lsak, demak biz allaqachon bilamiz

$$g_1(w_0), g_2(g_1(w_0)), ..., g_{m-1}(...g_1(w_0)...)$$

keyin biz quyidagicha harakat qilamiz:

- biz hosilani olamiz g_m nuqtada $g_{m-1}(...g_1(w_0)...)$;
- hosilaga ko'paytiring g_{m-1} nuqtada
- va hokazo, biz hosilaga yetgunimizcha g_1 nuqtada w_0

Biz buni rasmda tarozi bo'yicha differentsiatsiyani bosqichma-bosqich chizish orqali tasvirlaymiz w_i bitta ob'ektda logistik regressiyani yo'qotish funktsiyalari(ya'ni 1 o'lchamdagi Butch uchun):

Barcha omillarni birlashtirib, biz olamiz:

$$\frac{\partial f}{\partial w_1} = x_1(-y)\ \ e^{-y(w_0+w_1x_1+w_2x_2)}\ \ \frac{-1}{1+e^{-y(w_0+w_1x_1+w_2x_2)}}$$

$$\frac{\partial f}{\partial w_2} = x_2(-y)\ \ e^{-y(w_0+w_1x_1+w_2x_2)}\ \ \frac{-1}{1+e^{-y(w_0+w_1x_1+w_2x_2)}}$$

Shunday qilib, biz birinchi navbatda barcha oraliq qiymatlarni hisoblash uchun oldinga o'tish amalga oshirilganligini ko'ramiz (va ha, barcha oraliq ko'rinishlar xotirada saqlanishi kerak) va keyin barcha gradientlar bitta o'tishda hisoblanadigan backward pass boshlanadi.
Matritsani farqlashga bag'ishlangan paragrafda biz qisman hosilalarni alohida hisoblash yomon degan savolni ko'taramiz, matritsani hisoblash yaxshiroq. Ammo yana bir sabab bor: matritsa hosilasi bilan ham, printsipial jihatdan, siz har doim ham muomala qilishni xohlamaysiz. Oddiy misolni ko'rib chiqing. Aytaylik, X^r va X^{r+1} - ikkita ketma-ket oraliq vakillik $N \quad M$ va $N \quad K$, funktsiya bilan bog'liq $X^{r+1} = f^{r+1}(X^r)$. Aytaylik, biz qandaydir tarzda hosilani hisobladik

$$\frac{\partial L}{\partial X^r_{st}} = \sum_{ij} \frac{\partial f^{r+1}_{ij}}{\partial X^r_{st}} \frac{\partial L}{\partial X^{r+1}_{ij}}$$

N^2MK

x

A

$[D_{x_0} f](x - x_0) = \{$

$n \quad n$

$f(x) = x^x Ax, A \quad Mat_n R$

$$\frac{\partial f^{r+1}}{\partial X^r}$$

$NM + NK$

yo'qotish funktsiyalari , keyin

$$\frac{\partial L}{\partial X^r_{st}} = \sum_{ij} \frac{\partial f^{r+1}_{ij}}{\partial X^r_{st}} \frac{\partial L}{\partial X^{r+1}_{ij}}$$

Va biz ikkala gradient bo'lsa ham, buni ko'ramiz va $\frac{\partial L}{\partial X^r_{st}}$ ular shunchaki matritsalardir, hisoblash jarayonida "to'rt o'lchovli kub" paydo bo'ladi , hatto saqlash juda og'riqli: u juda ko'p xotirani talab qiladi (zararsiz bilan solishtirganda $NM + NK$, gradientlarni saqlash uchun zarur). Shuning uchun men *oraliq hosilalarni xohlayman* $\frac{\partial f^{r+1}}{\partial X^r}$ *hisoblangan ob'ektlar sifatida emas va o'zgartiradigan transformatsiyalar sifatida ichida* $\frac{\partial L}{\partial X^r_{st}}$ Quyidagi paragraflarning maqsadi aynan shunday bo'ladi: *ma'lum bir qatlamdan o'tayotganda Error backpropagation jarayonida gradient qanday o'zgarishini tushunish*.
O'zingizdan so'raysiz: endi matritsani farqlash bo'yicha darslikning paragrafini o'qib chiqishim kerakmi?

Qarshi savol. Funktsiyaning hosilasini vektor bo'yicha toping x :

$f(x) = x^x Ax, A \quad Mat_n R$ матрицаразмера— матрица размера $n \quad n$

Va hamma narsa qanday o'zgaradi, agar bu ham bog'liq x ? Funktsiya gradienti nima, agar skalarmi? Agar siz hozir qalam va qog'ozni olib, hamma narsani sanashga tayyor bo'lsangiz, unda matritsani farqlash haqida o'qishingiz shart emas. Ammo biz ushbu paragrafni ko'rib chiqishni maslahat beramiz, agar biz bundan keyin foydalanadigan belgilar sizga tushunarsiz bo'lib tuyulsa: matritsani farqlash uchun yagona belgi, afsuski, insoniyat hali ixtiro qilmagan va uni biridan ikkinchisiga o'tkazish har doim ham oson emas.

Biz darhol bizni qiziqtirgan narsaga o'tamiz: murakkab funktsiyalarning gradientlarini hisoblash.

Eslatib o'tamiz, murakkab funktsiyaning hosilasi formulasi quyidagicha:

Endi gradientlar bilan shug'ullanamiz. Ruxsat bering - skalar funktsiyasi. Keyin

Boshqa tomondan,

Ushbu formula xatoning teskari tarqalish mexanizmining yuragi hisoblanadi. U shunday deydi: agar biz qandaydir tarzda ba'zi oraliq ko'rinishlardan o'zgaruvchilar bo'yicha yo'qotish funktsiyasining gradientini olgan bo'lsak neyron tarmoq va shu bilan birga gradient qatlamdan o'tayotganda qanday o'zgarishini bilamiz o'rtasida va (ya'ni, ular orasidagi qatlamning differentsialiga konjugat qanday ko'rinishga ega), keyin biz darhol gradientni va o'zgaruvchilardan topamiz

Shunday qilib, qatlamma-qavat, biz hamma narsada gradientlarni hisoblaymiz birinchi qatlamlarga qadar.

Keyinchalik, ba'zi umumiy qatlamlardan o'tayotganda gradientlar qanday o'zgarishini aniq bilib olamiz.

Xato backpropagation algoritmini (xatoni teskari tarqatish algoritmi) tasvirlab, oldingi munozarani umumlashtiramiz. Aytaylik, bizda hozirgi tarozi qiymatlari bor va biz SGD-ning mini-Butch qadamini qo'ymoqchimiz . Biz quyidagilarni qilishimiz kerak:

1. Barcha oraliq ko'rinishlarni hisoblash va eslab qolish orqali oldinga o'tishni amalga oshiring .

2. Backward pass yordamida barcha gradientlarni hisoblang.

3. Olingan gradientlardan foydalanib, SGD qadamini bajaring.

Algoritmni skalyar chiqish bilan ikki qavatli neyron tarmog'i misolida tasvirlaymiz. Oddiylik uchun biz chiziqli qatlamlarda bo'sh atamalarni qoldiramiz.

Bunday holda, qatlam atrofida nima bo'layotganini umuman bilishi shart emas. Ya'ni, qatlam haqiqatan ham forward pass va backward pass-ni qanday qilishni biladigan alohida shaxs sifatida dasturlashtirilishi mumkin, shundan so'ng qatlamlar mexanik ravishda, konstruktordagi kublar singari, bitta birlik sifatida ishlay oladigan katta tarmoqqa yig'iladi.

Bundan tashqari, ko'p hollarda chuqur o'rganish kutubxonalari mualliflari sizga allaqachon g'amxo'rlik qilishgan va ifodalarni avtomatik ravishda farqlash vositalarini (autograd) yaratishgan. Shuning uchun, neyron tarmoqni dasturlash orqali siz deyarli har doim faqat oldinga o'tish, ma'lumotlarni to'g'ridan-to'g'ri o'zgartirish, kutubxonaga hamma narsani mustaqil ravishda farqlash imkonini berish haqida o'ylashingiz mumkin. Bu neyron tarmoq kodini juda tushunarli va ifodali qiladi (ha, aslida u ham katta va qo'rqinchli bo'lishi mumkin, ammo bo'sh vaqtingizda ba'zi bir eskirgan neyron tarmoq kodini va hal qiluvchi daraxtlardagi gradientni kuchaytirish kodini taqqoslang va farqni his eting). Xatoni orqaga qaytarish usuli sizga gradientlarni qulay hisoblash imkonini beradi, ammo keyin ular bilan biror narsa qilish kerak va yaxshi eski SGD zamonaviy tarmoqni o'rganishni uddalay olmaydi.

FOYDALANILGAN ADABIYOTLAR

1. Aizerman, M. A., Braverman, E. M., & Rozonoer, L. I. (1964). Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control, 25, 821–837.
2. Anlauf, J. K., & Biehl, M. (1989). The adatron: an adaptive perceptron algorithm. Europhysics Letters, 10(7), 687–692.
3. Block, H. D. (1962). The perceptron: A model for brain functioning. Reviews of Modern Physics, 34, 123–135. Reprinted in ”Neurocomputing” by Anderson and Rosenfeld.
4. Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, pp. 144–152.
5. Cesa-Bianchi, N., Freund, Y., Haussler, D., Helmbold, D. P., Schapire, R. E., & Warmuth, M. K. (1997). How to use expert advice. Journal of the Association for Computing Machinery, 44(3), 427–485.
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.
7. Friess, T., Cristianini, N., & Campbell, C. (1998). The kernel-adatron: A fast and simple learning procedure for support vector machines. In Machine Learning: Proceedings of the Fifteenth International Conference.
8. Gallant, S. I. (1986). Optimal linear discriminants. In Eighth International Conference on Pattern Recognition, pp. 849–852. IEEE.
9. Nurmamatovich, T. I. (2024, April). BIR QATLAMLI PERCEPTRONNI O ‘QITISH. In *" CANADA" INTERNATİONAL CONFERENCE ON DEVELOPMENTS İN EDUCATİON, SCİENCESAND HUMANİTİES* (Vol. 17, No. 1).
10. Nurmamatovich, T. I. (2024, April). SUN'IY NEYRONNING MATEMATIK MODELI HAMDA FAOLLASHTIRISH FUNKTSIYALARI. In *" USA" INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE TOPICAL ISSUES OF SCIENCE* (Vol. 17, No. 1).
11. Nurmamatovich, T. I. (2024, April). SUNIY NEYRON TORLARINI ADAPTIV KUCHAYTIRISH USULI. In *" USA" INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE TOPICAL ISSUES OF SCIENCE* (Vol. 17, No. 1).
12. Nurmamatovich, T. I. (2024, April). SUNIY NEYRON TORLARINI ADAPTIV KUCHAYTIRISH USULI. In *" USA" INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE TOPICAL ISSUES OF SCIENCE* (Vol. 17, No. 1).
13. Tojimamatov, I. N., Olimov, A. F., Khaydarova, O. T., & Tojiboyev, M. M. (2023). CREATING A DATA SCIENCE ROADMAP AND ANALYSIS. PEDAGOGICAL SCIENCES AND TEACHING METHODS, 2(23), 242-250.

14. Тожимаматов, И. Н. (2023). ЗАДАЧИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ. PEDAGOG, 6(4), 514-516.
15. Muqaddam, A., Shahzoda, A., Gulasal, T., & Isroil, T. (2023). NEYRON TARMOQLARDAN FOYDALANIB TASVIRLARNI ANIQLASH USULLARI. SUSTAINABILITY OF EDUCATION, SOCIO-ECONOMIC SCIENCE THEORY, 1(8), 63-74.
16. Raximov, Q. O., Tojimamatov, I. N., & Xo, H. R. O. G. L. (2023). SUNIY NEYRON TARMOQLARNI UMUMIY TASNIFI. Scientific progress, 4(5), 99-107.
17. Ortiqovich, Q. R., & Nurmamatovich, T. I. (2023). NEYRON TARMOQNI O ‘QITISH USULLARI VA ALGORITMLARI. Scientific Impulse, 1(10), 37-46.
18. Tojimamatov, I. N., Mamalatipov, O., Rahmatjonov, M., & Farhodjonov, S. (2023). NEYRON TARMOQLAR. Наука и инновация, 1(1), 4-12.
19. Tojimamatov, I. N., Mamalatipov, O. M., & Karimova, N. A. (2022). SUN’IY NEYRON TARMOQLARINI O ‘QITISH USULLARI. Oriental renaissance: Innovative, educational, natural and social sciences, 2(12), 191-203.
20. Muqaddam, A., Shahzoda, A., Gulasal, T., & Isroil, T. (2023). NEYRON TARMOQLARDAN FOYDALANIB TASVIRLARNI ANIQLASH USULLARI. SUSTAINABILITY OF EDUCATION, SOCIO-ECONOMIC SCIENCE THEORY, 1(8), 63-74.
21. Raximov, Q. O., Tojimamatov, I. N., & Xo, H. R. O. G. L. (2023). SUNIY NEYRON TARMOQLARNI UMUMIY TASNIFI. Scientific progress, 4(5), 99-107.
22. Raxmatjonova, M. N., & Tojimamatov, I. N. (2023). BIZNESDA SUNIY INTELEKT TEXNOLOGYALARI VA ULARNI AHAMIYATI. Лучшие интеллектуальные исследования, 11(3), 46-52.