

MALLAYEV OYBEK USMANKULOVICH
Texnika fanlari falsafa doktori (PhD), dotsent,
Perfect University, Toshkent, O'zbekiston,
mallayev@perfectuniversity.uz, <https://orcid.org/0000-0003-3966-1028>

Usarboyev Shaxzod Shavkatovich
magistrant, Alfraganus University,
Toshkent, O'zbekiston,
ushahzod96@gmail.com, <https://orcid.org/0000-0003-3966-1028>

Nurova Zaxro Baxrom qizi
magistrant, University of Management
and Future Technologies, Toshkent, O'zbekiston,
nurovazahro@gmail.com

MASHINAVIY O'QITISH ALGORITMLARIASOSIDA MAKTAB BITIRUVHILARINING YO'NALISHLARINI ANIQLASH MODEL



Annotatsiya. Ushbu maqolaning maqsadi mashinaviy o'qitish (machine learning) usullaridan foydalangan holda maktab o'quvchilarini fanlardan olgan baholari va softskills ko'nikmalari bo'yicha ularning yo'nalishlarini aniqlashdir. Maqolada ko'p o'zgaruvchili chiziqli regressiya yordamida sigmoid funksiya qurish murakkabliklari ko'rib chiqildi hamda o'quvchilarni 10 yil davomida ta'lim yo'nalishi bo'yicha tanlangan fanlardan olgan baholari va ularni turli parametrlari, sabablari hamda o'quvchining imkoniyatlari raqamlashtirildi. Ushu raqamlar yordamida trening ma'lumotlar top'lami tashkil qilindi. Natijada maktab o'quvchilarini 10 yil davomida o'qigan fanlari va ulardan olgan baholarining klassifikatsiyasi ishlab chiqildi. Neyron tarmoq arxitekturalari, modullari, mashinaviy o'qitish algoritmlarida eng ko'p qo'llanilayotgan aktivlashtirish funksiyalari, o'qitish usullari hamda chiziqli va logistik regressiya qurish usullari, kamchiliklari va imkoniyatlari tahlil qilingan. Ko'p o'zgaruvchili chiziqli regressiya uchun gradiyent tusish funksiyasini vektorlangan hisoblash orqali qulaylashtirish yo'llari o'rganib chiqildi. Chiziqli regressiyaning bu turida juda ko'p o'zgaruvchilar qatnashganligi uchun vektor hisob-kitoblar ancha qulayli isbotlangan. Vektor hisoblashlar yordamida gradiyent tushish jarayonlarini parallel hisoblash yo'llari ham ko'rib o'tilgan. Hususan, trening ma'lumotlar jadvalining ustunlarini qo'shish, koeffitsiyentlarni transpozitsiyalash - AT, chiziqli funksiyaning vektorlangan ko'rinishlari, gradiyent tushish uchun giperparametrlar (o'rganish darajasi - , qadamlar soni) aniqlab olindi.

Annotation. The purpose of this article is to determine the orientation of schoolchildren in terms of grades in subjects and soft skills using machine learning methods. The article examined the difficulties of constructing a sigmoid function using multivariate linear regression, and also digitized grades obtained in selected subjects in the field of student education over 10 years and their various parameters, reasons and student capabilities. . Using these numbers, a training data set was created. As a result, a classification of subjects studied by schoolchildren over 10 years and their assessments was developed. Neural network architectures, modules, the most commonly used activation functions in machine learning algorithms, training methods and methods for constructing linear and logistic regression, disadvantages and opportunities are analyzed. Ways to simplify the gradient descent function for multivariate linear regression by vector calculation have been studied. Because there are many variables involved in this type of linear regression, vector calculations have proven to be more convenient. Methods for parallel calculation of gradient descent processes using vector calculations are also considered. In particular, the addition of training data table columns, transposition of coefficients - AT, vectorized representation of a linear function, hyperparameters for gradient descent (learning rate - , number of steps) were defined.

Аннотация. Цель данной статьи – определить ориентацию школьников с точки зрения оценок по предметам и softskills с помощью методов машинного обучения. В статье были рассмотрены сложности построения сигмовидной функции с использованием многомерной линейной регрессии, а также оцифрованы оценки, полученные по выбранным предметам в области образования студентов за 10 лет и их различные параметры, причины и возможности студента. Используя эти числа, был создан набор обучающих данных. В результате была разработана классификация предметов, изучаемых школьниками за 10 лет, и их оценки. Проанализированы архитектуры нейронных сетей, модули, наиболее часто используемые функции активации в алгоритмах машинного обучения, методы обучения и методы построения линейной и логистической регрессии, недостатки и возможности. Изучены пути упрощения функции градиентного спуска для многомерной линейной регрессии путем векторного расчета. Поскольку в этом типе линейной регрессии участвует множество переменных, векторные вычисления оказались более удобными. Также рассмотрены способы параллельного расчета процессов градиентного спуска с использованием векторных вычислений. В частности, были определены добавление столбцов таблицы обучающих данных, транспонирование коэффициентов - AT, векторизованное представление линейной функции, гиперпараметры для градиентного спуска (скорость обучения - , количество шагов).

KIRISH

Joriy yil mutaxassislar mahalliy ta'lim tizimining kompleks tahlilini o'tkazish natijasida oliy ta'lim tizimida quyidagi kamchiliklar qayd etilgan:

Oliy ta'lim jarayonida nazariya va amaliyot yaxlitligi ta'minlanmaganligi. Talabalarning malakaviy amaliyotlarini ishlab chiqarish korxonalarida o'tkazish samarali tashkil etilmagani oqibatida bitiruvchilarning aksariyat qismi tayyor mutaxassis bo'lib chiqish o'rniga, ishga joylashgandan keyin qaytadan o'z kasbini, mutaxassisligini o'rganayotganligi, ta'lim sifatini nazorat qilish mexanizmi zamonaviy talablarga javob bermasligi, ta'lim muassasalarida malakali pedagog va boshqaruv kadrlarining yetishmasligi,

xorijiy ta'lim muassasalari bilan samarali hamkorlik yetarlicha yo'lga qo'yilmaganligi, iqtisodiyot sohasining istiqbolda kadrlar tayyorlashga buyurtmalarini shakllantirish, bitiruvchilarga qo'yiladigan malakaviy talablarni ishlab chiqish, tarmoqqa zarur bo'lgan mutaxassislarni tayyorlash sifatini ta'minlash jarayonidagi ishtiroki talabga javob bermaydi, oliy ta'lim – ilm-fan – ishlab chiqarish o'rtasida uzilishlar mavjud, lekin integratsiya ta'minlanmaganligi va shu kabi kamchiliklar aniqlangan. Bu kabi kamchiliklarni bartaraf etish mexanizmlarini ishlab chiqish bo'yicha bir qancha ishlar jadallik bilan olib borilmoqda. Lekin masalaning ikkinchi tarafi ham bor.

Ta'lim olish uchun kelayotgan talabalarning aksariyat foizi tanlangan yo'nalishida adashganini, yoki bo'lmasam ixtidori va qiziqishi yo'qligi takidlayotgani ayni haqiqat. Bunday vaziyatda ta'lim berishning eng zamonaviy usullarini qo'llagan holda darslarni tashkil qilsangiz ham mukammal amaliyot o'tash mexanizmlarini ishlab chiqsangiz ham bitirib chiqayotgan kadrlarning sifati ko'tarilmaydi va ko'tarilmayotganini ko'rish mumkin.

Bu masalning tub mohiyatidan kelib chiqib, maqolada oliy ta'lim muassasalariga o'qishga keladigan ilk, ixtidori va shu yo'nalishga qiziqishi yuqori talabalarni va ularni sonini mashinali o'qitish algoritmlari

yordamida avvaldan aniqlash algoritmi, sun'iy intellekt moduli va dasturiy vositasi taklif qilinadi.

Bu masalani yechishning sun'iy intellekt modulini ishlab chiqish uchun bir nech katta muammoli vaziyatlar mavjud. Ya'ni ishlab chiqish ko'p vaqt sarf qiladigan parametrlar. Bular:

- maktab o'quvchilarini baholash mexanizmlarini o'rganish;
- o'quvchilar fanlardan olayotgan baholari yo'nalishlar kesimida tahlil qilish;
- yo'nalishlar kesimida o'quvchilarni SoftSkills ko'nikmalarini aniqlash;
- yo'nalishlar uchun muhim fanlar ro'yxatini shakillantirish;
- har bir talabani ushbu muhim fanlardan olgan baholar jadvallari shakillantirish;

- yosh psixologiyasi, ixtidorlari, yutuqlarini inobatga olgan holda jadvallarni shakillantirish;

- taklif etilayotgan sun'iy intellekt modulini haqiqatga yaqin natijalar berishi uchun yillar kesimida fanlardan olgan baholar jadvallari shakillantirish.

Shu kabi shakillantirilishi uzoq muddatni talab etadigan, neyron tarmoqni o'qitish uchun zarur bo'lgan ma'lumotlar va ularning turli xil ko'rinishdagi parametrlarini aniqlash taklif qilinaotgan modul uchun juda ham zarur hisoblanadi.

O'quvchilarning bilim saviyasi va o'zlashtirish darajasi joriy nazorat va oraliq nazorat turlarini o'tkazish orqali baholanadi. Joriy nazorat — so'rovlar, nazorat ishlari yoki testlar tarzida o'quvchilarning bilimlari, ko'nikma va malakalari muntazam nazorat qilinadi. Oraliq nazorat — chorak tamom bo'lganda va o'quv dasturining tegishli bo'limi tugallangandan keyin o'quvchilarning bilimlari, ko'nikma va malakalarini baholash uchun amalga oshiriladi. U yozma nazorat ishi yoki testlar shaklida o'tkaziladi. Ushbu nazoratda foydalaniladigan materiallar o'quv fani o'qituvchisi tomonidan tayyorlanadi.

Har bir nazorat turi qanday shaklda o'tkazilishidan qat'i nazar, besh ballik («5», «4», «3», «2», «1») usulda butun sonlar yordamida baholanadi. O'quvchilarning nazorat turlari bo'yicha olgan ballari hamda choraklik (yarim yillik) va yillik (yakuniy) baholari, shuningdek, yillik reyting ko'rsatkich ballari sinf jurnaliga qayd etiladi. O'quvchining choraklik (yarim yillik), yillik (yakuniy) baholari va yillik reyting ko'rsatkich ballari uning tabelida qayd

etiladi.

Chorak, yillik va yakuniy baholash mexanizmlari:

Haftasiga 1 soatdan ortiq o'qitiladigan o'quv fanlari bo'yicha:

- chorak davomida joriy va oraliq nazoratlarda olingan ballar asosida choraklik baho;
- choraklik baholar va bosqichli nazoratda olingan ball asosida yillik baho aniqlanadi.

Haftasiga 0,5 va 1 soat o'qitiladigan o'quv fanlari bo'yicha:

- yarim yil davomida joriy va oraliq nazoratlarda olingan ballar asosida yarim yillik baho;
- yarim yillik baholar va bosqichli nazoratda olingan ball asosida yillik baho aniqlanadi.

Bosqichli nazorat o'tkazilmaydigan o'quv fanlari bo'yicha yillik baholar yarim oraliq va yillik baholarga qarab belgilanadi.

Bitiruvchilar uchun choraklik (yarim yillik) baholar asosida yillik baho, shuningdek yakuniy nazoratda olingan ball va yillik baho asosida yakuniy baho aniqlanadi.

Yakuniy davlat attestatsiyasi belgilanmagan o'quv fanlari bo'yicha yillik baho yakuniy baho hisoblanadi. Yakuniy baholar bitiruvchiga beriladigan umumiy o'rta ta'lim to'g'risidagi shahodatnomaga qayd etiladi.

Texnika yo'nalishi uchun muhim fanlar va ularni rivojlantirish parametrlari

Texnika yo'nalishida o'qishga

1- jadval. Trening ma'lumotlari

№	O'quv yillari										Faolligi	Yutuqlari	Natija
	1	2	3	4	5	6	7	8	9	10			
1	5	5	5	5	5	5	5	5	5	5	0	0	1
2	5	5	5	5	5	5	5	5	5	5	1	0	1
3	5	5	5	5	5	5	5	5	5	5	2	0	1
4	5	5	5	5	5	5	5	5	5	5	3	0	1
5	5	5	5	5	5	5	5	5	5	5	4	0	1
6	5	5	5	5	5	5	5	5	5	5	5	0	1
7	5	5	5	5	5	5	5	5	5	5	5	1	1
8	5	5	5	5	5	5	5	5	5	5	5	2	1
9	5	5	5	5	5	5	5	5	5	5	5	3	1
10	5	5	5	5	5	5	5	5	5	5	5	4	1
11	5	5	5	5	5	5	5	5	5	5	5	5	1
12	4	4	4	4	4	4	4	4	4	4	0	0	1
13	4	4	4	4	4	4	4	4	4	4	1	0	1
14	4	4	4	4	4	4	4	4	4	4	2	0	1
15	4	4	4	4	4	4	4	4	4	4	3	0	1
16	4	4	4	4	4	4	4	4	4	4	4	0	1
17	4	4	4	4	4	4	4	4	4	4	5	0	1
18	4	4	4	4	4	4	4	4	4	4	0	1	1
19	4	4	4	4	4	4	4	4	4	4	0	2	1
20	4	4	4	4	4	4	4	4	4	4	0	3	1
21	4	4	4	4	4	4	4	4	4	4	0	4	1
22	4	4	4	4	4	4	4	4	4	4	0	5	1
23	3	3	3	3	3	3	3	3	3	3	5	5	1
24	3	3	3	3	3	3	3	3	3	3	4	5	0
25	3	3	3	3	3	3	3	3	3	3	3	5	0
26	3	3	3	3	3	3	3	3	3	3	2	5	0
27	3	3	3	3	3	3	3	3	3	3	1	5	0
28	3	3	3	3	3	3	3	3	3	3	5	1	0
29	3	3	3	3	3	3	3	3	3	3	5	2	0
30	3	3	3	3	3	3	3	3	3	3	5	3	0
31	3	3	3	3	3	3	3	3	3	3	5	4	0
32	3	3	3	3	3	3	3	3	3	3	5	0	0

kirmoqchi bo'lgan o'quvchi asosan quyidagi fanlardan yaxshi baholarga o'qishi kerak:

- matematika;
- fizika;

Ikki fanidan bitta o'quvchini 10 yil davomida olgan baholar jadvali shakllantirildi

(1-jadval). Baholarning yillar kesimida o'zgarish parametrlari (variantlar) aniqlanib, katta hajimli jadvallar shakllantirildi. Bu qisqa ma'noda shuni anglatadiki, mashinali o'qitish algoritmlarida kiruvchi va trening ma'lumotlarini tashkil qiladi. Bunday ko'p parametrlar trening ma'lumotlarini qancha ko'p ko'paytirish modulning aniq ishlashiga va xatoliklarni kamayishiga olib keladi. Misol tariqasida matematika fanidan qo'yilgan baholar jadvali shakllantirildi hamda unga "faollik" va "yutuqlar" parametrlari qo'shildi. Chunki bu parametrlar yo'nalish uchun muhim parameter qiymatlarini beradi.

Taklif qilinayotgan mashinaviy o'qitish modelini 1- jadval asosidagi ma'lumotlar bilan to'ldirilgan matritsa bilan o'qitib olindi.

Ushbu jadvalda bitta o'quvchini 10 yil davomida matematika fanidan olgan baholar yillar kesimida berilgan. Faolligi, Yutuqlari va Natijalari ustunlari sun'iy intellekt modulini o'qitish va xatoliklarni minimallashtirish ya'ni o'quvchini yo'nalishini aniqlash noaniqligini kamaytirish uchun yuqorida takidlangan kriteriyalar asosida kiritildi. Taklif qilingan modul qaror qabul qilish uchun minimal trening ma'lumotlar jadvali (1- jadval) 32 qator yetarli bo'ldi. Aslida jadval 384 ta qatordan iborat. O'quvchini 1 fanidan 10 yil davomida olgan baholarining turli xil variantlari bilan boyitilgan. Kiruvchi har ma'lumotlar terning ma'lumotlar jadvalida mavjud bo'lmasa, modul dasturi natijasi chiqqandan keyin trening ma'lumotlar jadvaliga qo'shadi.

Maqolada taklif qilingan sun'iy intellekt modulini qurishda mashinaviy o'qitish turlaridan biri nazoratli o'qitish (Supervised Learning) dan foydalanildi.

Mashinaviy o'qitishda muhim qism, kompyuterlarga ma'lumotlarni o'rgatish uchun algoritmlar yaratish va ulardagi barcha o'zgarishlarni kompyuter o'z-o'zini o'rgatib borishidir. Ma'lumotlar to'plamini tahlil qilish, mos keladigan algoritmani tanlash, unga ma'lumotlarni o'rganish uchun foydalanish va natijalarni tahlil qilish mashinaviy o'qitishda asosiy elementlardir [10]. Bu soha quyidagi yirik yo'nalishlarni o'z

ichiga oladi:

Nazoratli o'qitish (Supervised Learning): Bu usulda, tizimga ta'lim maqsadida yaratilgan ma'lumotlar beriladi va tizim bu ma'lumotlarni o'rganadi. Bu o'qitish usuli o'rganish natijalarini bashorat qilishda va boshqa dasturlar ishlatilishida juda ko'p maqsadlarda qo'llaniladi. Ushbu turdagi o'qitish hozir ko'plab masalalarni yechishda qo'llanilmoqda. Trening ma'lumotlardan foydalangan holda kompyuter dasturi o'qitiladi (traning). Nazoratli o'qitish ham 2 ga bo'linadi:

- 1- Regressiya (Regressiya);
- 2- Klassifikatsiyalash (Tasniflash - classification).

Maktab bitiruvchilarining ta'lim yo'nalishlarini aniqlash modeli qurish uchun Regressiya o'qitish turi tanlangan.

Nazoratsiz o'qitish (Unsupervised Learning): Bu usulda tizimga ma'lumotlar beriladi, lekin ularning narxi (qiymati) aytilmagan. Tizim, ma'lumotlarni tahlil qilib, o'rganadi va ularda o'zgartirishlar qiladi. Usulning maqsadi belgilanmagan ma'lumotlar asosida modelni o'rgatish hisoblanadi.

Qisman nazoratli o'qitish (Semi-supervised Learning): Bu usulda nazoratsiz va nazoratli o'qitishning bir qismi jamlanadi. Bu usul ma'lumotlar qismini nazoratli, qolgan qismini esa nazoratsiz qilib o'rganishga ruxsat beradi.

Kuchaytirilgan o'qitish (Reinforcement Learning): Bu usulda tizim o'zini o'zgartirishlarni o'rganadi, kelajakdagi harakatlarni aniqlaydi va natijalarning yaxshi bo'lishi uchun o'zini o'zgartiradi.

Model uchun tanlangan chiziqli regressiya mashinaviy o'qitish sohasining eng boshlang'ich usullaridan biri bo'lib, uning yordami bilan davomiy bo'lgan qiymatlar prognoz qilinadi.

Chuqur o'qitish texnologiyalari

So'nggi paytlarda chuqur o'qitish texnologiyalari har bir sohada o'z imkoniyatlarini kengaytirib bormoqda. Masalan, turli xil murakkab muammolarini hal qilishda keng qo'llanilmoqda. Bunga hisoblash quvvatining oshishi va ko'p sonli jadval ko'nishidagi ma'lumotlar, shuningdek, bulutli saqlashdan foydalanish imkoniyati yordam beradi [4].

Neyron tarmoqlar ta'limda quyidagilarda qo'llanilmoqda:

- murakkab hisoblash jarayonlarida;
- avvaldan prognozlash sifat ehtimolini oshirishda;
- dars davomatlarini sifati,

haqqoniyligi va online monitoringini amalga oshirishda;

- ma'lumotlarni tez qidirish va sifatli ma'lumotlarni topishda.

Fanlardan olingan baholar joylashgan, katta hajimli jadvallar va ulardagi parametrlar asosida turli murakkab muammolarni aniqlashda neyron tarmoq texnologiyalarining ro'li juda ham katta. Ular oliy ta'lim tizimida fan kontentlarini tayyorlash sifatini oshirish, ko'p yillar katta tajribaga ega o'qituvchilarning bilimlaridan turli sun'iy o'qituvchilarni yartish hamda ulardan bir vaqtning o'zida bir nechta joylarda foydalanish imkoniyatlarini beradi. Bu sun'iy neyron tarmoqlardan foydalanish tufayli yaratilayotgan ulkan imkoniyatlarning faqat kichik bir qismidir.

Chuqur neyron tarmoqlar ta'lim tizimida yillar davomida hal qilinishi kerak bo'lgan muammolarni soniyalarda hal qiladi [1][2][3].

Ularning yordami bilan ta'lim muassasalarida darslar sifatini, talabalarni bilim olish tendensiyalarini aniqlash va turli xil ta'limiy vaziyatlar va boshqa ko'plab muammolarni aniqlash mumkin [7].

Sun'iy neyron tarmoqlardan foydalanishning asosiy muammolaridan biri bu - loyihalashtirilgan tarmoqning arxitekturasini ham, natijalarning talab qilinadigan aniqlik darajasi uchun zarur bo'lgan murakkabligini ham oldindan aniqlashning mumkin emasligidir. Agar talabaning bilim darajasi juda yuqori bo'lsa, arxitekturani yanada murakkablashtirishga to'g'ri keladi.

Faqat bitta yashirin qatlama ega neyron tarmoqlar eng oddiy vazifalarni bajarishi mumkin. Murakkabroq muammolarni hal qilish uchun zarur bo'lgan yashirin qatlamlar sonining ko'payishi hisoblashga sarflanadigan quvvatini va vaqtni oshirishga olib keladi.

Kompyuter protsessorlarining quvvatini oshirish va klasterlash tizimlarini takomillashtirish orqali ushbu muammoni hal qilish uchun ko'plab ilmiy tadqiqotlar olib borilmoqda [8].

Taklif qilingan modulning barcha ma'lumotlar jadvali ikki qismga bo'lingan: o'qitish va sinov.

Har bir jadvalning "Natija" ustuni (1) mavjud va (0) mavjud emas bo'lgan fanlardan olinga baholardan iborat. Barcha 100% ma'lumotlar quyidagicha bo'linadi: umumiy ma'lumotlar to'plamining 80% modelni o'rgatish uchun, qolgan 20% sinov uchun ajratilgan. Quyida bir nechta neyron tarmoqlarning ishlashini batafsil

ko'rib chiqilgan (modellash Pythonida Sklearn kutubxonasi yordamida amalga oshiriladi).

Neyron tarmoqning asosiy afzalliklaridan biri shundaki, u juda murakkab funksiyalarni ifodalay oladi.

Agar faollashtirish funktsiyalari qo'llanilmasa, chiqish signali bir darajali polinom bo'lgan chiziqli funktsiya bo'ladi. Chiziqli tenglamalarni yechish oson bo'lsada, ular cheklangan murakkablik omiliga ega va shuning uchun ma'lumotlardan murakkab funktsional xaritalarni o'rganish qobiliyati kamroq. Shunday qilib, faollashtirish funktsiyasiz neyron tarmoq cheklangan imkoniyatlarga ega chiziqli regressiya modeli bo'ladi.

Eng mashhur faollashtirish funktsiyalari quyidagilarni o'z ichiga oladi:

oladi:

- **sigmoid** $\sigma = \frac{1}{1 + e^{-x}}$

funktsiya haqiqiy sonni oladi va uni 0 dan 1 gacha bo'lgan songa aylantiradi. Xususan, katta manfiy sonlar 0 ga, katta musbat sonlar esa 1 ga aylanadi.

- **Tanh funksiyasi** $x = \frac{2}{1 + e^{-2x}} - 1$

haqiqiy sonni [-1, 1] diapazoniga siqib chiqaradi. Uning chiqishi nol markazga ega. Bu sozlangan sigmasimon tanh funktsiyasi ekanligini ko'rish mumkin.

- **Softmax** funktsiya 0 va 1 qiymatlari o'rtasida o'zgarib turadigan natijani hosil qiladi va ehtimolliklar yig'indisi 1 ga teng.

- **ReLU (Rectified linear unit)** U $(x) = \max(0, x)$ ffunktsiyasini baholaydi. Boshqacha qilib aytganda, faollashtirish oddiygina nol chegaraga ega.

- **ELU (Exponential Linear Units)**

$$= \begin{cases} a(e - 1), & x \leq 0 \\ x, & x > 0 \end{cases}$$

x bu yerda manfiy komponent eksponensial belgilar yordamida modellashtiriladi.

Chuqur o'rganish, ayniqsa, yuzni aniqlash, matnni tarjima qilish va ovozni aniqlash kabi murakkab identifikatsiya dasturlari uchun juda mos keladi[5][6].

Neyron tarmog'ining natijalari ko'p jihatdan parametrlarni tanlashga bog'liq. Ularni aniqlashning yagona umumiy algoritmi yo'q, shuning uchun hamma narsa sinov va xatoliklar orqali amalga

oshiriladi. Eng maqbul qiymatlarni aniqlash uchun turli qiymatlarga ega bo'lgan bir nechta turli xil variantlarni amalga oshirish kerak.

Modelning matematik asoslari:

Model uchun chiziqli funktsiya formulasi:

$$Y = xA^T + b \quad (1)$$

Chiziqli regressiya bilan ishlaganda ma'lumotlarni ifoda etuvchi eng yaxshi regressiya chizig'ini topish kerak bo'ladi. Regeressiya chizig'i (1) bilan hisoblanganda XT hisoblanadi. Bunda ushbu konstanta orqali prognoz qilingan natija ma'lumotlar to'plamida keltirilgan haqiqiy qiymatga imkon qadar yaqin bo'lishi kerak. Ya'ni prognoz qilingan qiymat va haqiqiy qiymat o'rtasidagi farq juda ham kam bo'lishi kerak. Ushbu farq xatolik deyiladi. Xatolikni aniqlash uchun o'rtacha kvadratik xatolik funktsiyasi tanlanadi.

$$j = \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 \quad (2)$$

Regressiya koeffitsientlarini yangilash orqali xatolikni kamaytirish mumkin. Shuning uchun (2) bilan optimallashtirish amalga oshiriladi. Modulning natijasi traning ma'lumotlar jadvali (1- jadval)ning bir nechta ustunlariga bog'liqligi uchun ko'p o'zgaruvchili chiziqli regressiyani hisoblash bosqichlari amalga oshiriladi.

Buni amalga oshirish uchun bazi belgilashlar kiritilgan:

m – ma'lumotlar to'plami soni (qatorlar soni);

n– o'zgaruvchilar soni (ustunlar soni);

$x^{(i)}$ – i qatordagi o'zgaruvchilar;
 $x_{-j}^{(i)}$ – i qatordagi j- o'zgaruvchining qiymati;

X – ma'lumotlar to'plamidagi o'zgaruvchilarni ifoda etuvchi matritsa.

a_j – j- koeffitsient;
A – barcha koeffisientlarni ifodalovchi matritsa yoki vektor. Uning o'lchami $(n \times 1)$;

$y^{(i)}$ – i- qatordagi prognoz qilinadigan belgining ma'lumotlar to'plamidagi asl qiymati;

Y – prognoz qilingan barcha $y^{(i)}$ larni ifodalovchi birlik matritsa yoki vector. Uning o'lchami $(1 \times m)$;

$f(x)$ – prognozni hisblash formulasi;
Jumladan, 1- jadvalda keltirilgan ma'lumotlar to'plamida:

m= 32, ya'ni ma'lumotlar to'plami soni (qatorlar soni) 32 ga teng;

n = 12, o'zgaruvchilar soni (ustunlar soni) 12 ga teng. Bunda o'zgaruvchilarni

O'quv yillari (10 ta ustun), Faolliqi, Yutuqlari ustunlari tashkil etadi;

$x^{(1)} = [5, 5, 5, 5, 5, 5, 5, 5, 5, 0, 0, 1]$ ya'ni 1- qatorda keltirilgan **O'quv yillari (10 ta ustun), Faolliqi, Yutuqlari** ustunlarining qiymatlar to'plami;

$x_{11}^1 = 0$, ya'ni 1- qatordagi 11- o'zgaruvchining (**Faolliqi**) qiymati;

$y^{(1)} = 1$, ya'ni 1- qatorda prognoz qilinadigan belgining ma'lumotlar to'plamidagi asl qiymati;

α – o'rganish darajasi (learning rate);

davr – gradiyet tushishdagi qadamlar soni (epochs);

1- jadvaldagi ma'lumotlar to'plami-ga ko'p o'zgaruvchili chiziqli regressiya qo'llanilsa, (1) formula quyidagi ko'rinishni oladi:

$$f(x) = a_0 + a_1 \times x_1 + a_2 \times x_2 + a_3 \times x_3 + \dots + a_n \times x_n \quad (3)$$

Ko'p o'zgaruvchili chiziqli regressiya uchun gradiyent tusish funktsiyasini vektorlangan hisoblash orqali qulaylashtirildi [9]. Chiziqli regressiyaning bu turida juda ko'p o'zgaruvchilar qatnashganligi uchun vektor hisob-kitoblar ancha qulay hisoblanadi. Shning uchun vektor hisoblarda qulaylik bo'lishi uchun (3) ga o'zgarishlar kiritiladi:

$$f(x) = a_0 \times x_0 + a_1 \times x_1 + a_2 \times x_2 + \dots + a_n \times x_n \quad (4)$$

4- formulani ma'lum bir ma'lumotlar to'plami qatoriga moslashtirilsa, quyidagi tenglik kelib chiqadi:

$$\begin{aligned} \hat{y}^1 &= a_0 \times x_0^1 + a_1 \times x_1^1 + a_2 \times x_2^1 + \dots + a_n \times x_n^1, \\ \hat{y}^2 &= a_0 \times x_0^2 + a_1 \times x_1^2 + a_2 \times x_2^2 + \dots + a_n \times x_n^2, \\ \hat{y}^3 &= a_0 \times x_0^3 + a_1 \times x_1^3 + a_2 \times x_2^3 + \dots + a_n \times x_n^3, \\ &\vdots \\ \hat{y}^m &= a_0 \times x_0^m + a_1 \times x_1^m + a_2 \times x_2^m + \dots + a_n \times x_n^m, \end{aligned} \quad (5)$$

5- formulani ma'lumotlar jadvali qatorlarining har biri uchun tadbiq qilindi, bunda - 1- qatorning prognoz qilingan qiymati, y^m - m- qatorning prognoz qilingan qiymati. Shu formula elementlarini matritsa yordamida ifodalaniladi. Prognozlar matritsasi:

$$\hat{y} = \begin{bmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \dots \\ \hat{y}^{(m)} \end{bmatrix} \quad (6)$$

5- formulada $(m \times 1)$ bo'lgan y matritsa hosil bo'ldi. O'zgaruvchilar matritsasa quyidagicha ko'rinishda bo'ladi:

$$X = \begin{bmatrix} x_0^1, x_1^1, \dots, x_n^1 \\ x_0^2, x_1^2, \dots, x_n^2 \\ \dots \\ x_0^m, x_1^m, \dots, x_n^m \end{bmatrix} \quad (7)$$

$(m \times (n+1))$ o'lchamli matritsa hosil bo'ldi. Bunda m - qatorlar soni, $(n+1)$ - o'zgaruvchilar soni. Quyidagi formula koeffitsiyentlar matritsasini anglatadi:

$$A = [a_0, a_1, \dots, a_n] \quad (8)$$

Vektor ko'paytirishlar uchun A vektorni transpozitsiyasi (qatorni ustunlar bilan almashtirish) qilinadi va quyidagi ko'rinishda ifolaniladi:

$$A^T = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{bmatrix} \quad (9)$$

Natijada o'lchami $((n+1) \times 1)$ bo'lgan matritsa, ya'ni vektor hosil bo'ladi. Keyingi qadamda X matritsani AT vektorga ko'paytirish mumkin.

$$\begin{bmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \dots \\ \hat{y}^{(m)} \end{bmatrix} = \begin{bmatrix} x_0^1, x_1^1, \dots, x_n^1 \\ x_0^2, x_1^2, \dots, x_n^2 \\ \dots \\ x_0^m, x_1^m, \dots, x_n^m \end{bmatrix} \times \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{bmatrix} \quad (10)$$

Ko'p o'zgaruvchili chiziqli regressiyaning formulasini 10- formulada keltirilgan matritsalar orqali ifodalanildi. Uning qisqa ko'rinishi quyidagicha:

$$\hat{y} = X \times A^T \quad (11)$$

11- formulada, y - o'lchami $(m \times 1)$ bo'lgan vektor, X - o'lchami $(m \times (n+1))$ bo'lgan matritsa hamda A^T - o'lchami $((n+1) \times 1)$ bo'lgan vektordir.

Gradiyent tushish fuksiyasi yordamida qiymat funksiyasini minimallashtiruvchi koeffitsiyentlarni topish uchun ko'p o'zgaruvchili chiziqli regressiyaning qiymat funksiyasi aniqlash, hisoblashlarga yanada qulay bo'lishi uchun 2- formula quyidagi ko'rinishi aniqlab olindi:

$$J = \frac{1}{2m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)})^2 \quad (12)$$

Gradiyent tushishning asosiy maqsadi qiymat funksiyasini minimallashtirish bo'lgani uchun yangi koeffitsiyent umumiy maqsadni o'zgartirmaydi.

$$\min \frac{1}{m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)})^2 \Leftrightarrow \min \frac{1}{2m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)})^2 \quad (13)$$

13-formula yordamida qiymat funksiya aniqlab olindi. Uni minimallashtirish uchun graadiyent tushish formulalaridan foydalanamiz. Bir o'zgaruvchili regressiyadan ma'lumki, funksiyani minimallashtirish uchun ushbu funksiyaning eng minimal qiymatiga olib boruvchi koeffitsiyentlarini topish kerak. Qiymat funksiya koeffitsiyentlar yordamid ifodalanilsa, quyida ko'rinishga keladi:

$$J = \frac{1}{2m} \sum_{i=1}^m ((a_0 \times x_0^{(i)} + a_1 \times x_1^{(i)} + a_2 \times x_2^{(i)} + \dots + a_n \times x_n^{(i)}) - y^{(i)})^2 \quad (14)$$

14-formuladagi qiymat funksiyasida $(n+1)$ ta koeffitsiyent qatnashadi. Bularni quyidagicha yozish mumkin:

$$A = [a_0, a_1, \dots, a_n] \quad (15)$$

Qiymat funksiyani eng minimal qiymatini topish uchun 15- formulada berilgan koeffitsiyentlarni optimal qiymatini topish kerak. Buning uchun qiymat funksiyasidan har bir koeffitsiyentga nisbatan hosila olinadi. Olingan hosilalardan foydalanib, ko'p o'zgaruvchili chiziqli regressiyaning gradiyent tushish formulasini hosil qilamiz:

$$\begin{cases} a_0 = a_0 - \alpha \frac{1}{m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)}) \times x_0^{(i)} \\ a_1 = a_1 - \alpha \frac{1}{m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)}) \times x_1^{(i)} \\ a_2 = a_2 - \alpha \frac{1}{m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)}) \times x_2^{(i)} \\ \dots \\ a_n = a_n - \alpha \frac{1}{m} \sum_{i=1}^m (f(x^{(i)}) - y^{(i)}) \times x_n^{(i)} \end{cases} \quad (16)$$

16-formulada dastlab koeffitsiyentlarning qiymatlariga tasodifiy sonlar (- o'rganish darajasi) beriladi keyin qadamda amallar bajariladi. Ushbu amallar koeffitsiyentlarning optimal qiymati topilguncha qaytariladi (davr tugaguncha). Buning uchun davrning har bir qadamida qiymat funksiya hisoblab boriladi. Ushbu amallardagi koeffitsiyentlarning qiymatlari parallel ravishda yangilanadi.

Model berilganlar ni bir nechta qatlamdan o'tkazadi. Dastlab ko'p

o'zgaruvchili chiziqli regressiya hisoblanadi. Chunki 1- jadvalda keltirilgan trening ma'lumotlari jadvalining "Faolligi" va "Yutuqlari" ustuni qiymatlari 5 dan yuqori. Chiziqli regressiyadan maqsad doimiy nishonni toppish, Leogistik regressiyada esa diskret (ma'lum bir turdagi) nishonlarni prognoz qiladi.

Berilgan trening ma'lumotlar to'plami diskret qiymatli ma'lumotlar to'plamiga tegishli bo'lganligi uchun taklif qilinayotgan modelda sigmoid funksiyadan foydalnildi.

Sigmoid funkjsiya logistic funksiya deb ham ataladi. Uning asosiy vazifasi berilgan sonni (0,1) oraligdagi qiymatini aniqlashdir. Mashinaviy o'qitishda esa sigmoid funksiya prognoz qilgan qiymatni ehtimolini aniqlashda foydalaniladi. Uni formula 17- formulada berilgan.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (17)$$

Bunda z - sigmoid funksiyaga kiruvchi o'zgaruvchi. Sigmoid funksiyaning logistic regressiyaga tadbqiq quyidagicha:

$$f(x) = XA^T \quad (18)$$

Bunda ikkita matritsa ko'paytmasi (X va A^T) davomli, ya'ni chiziqli qiymatlarni beradi. Undan diskret qiymatlarni olish uchun sigmoid funksiyani tadbqiq qilish kerak.

$$\sigma(Z) = \sigma(XA^T) \quad (19)$$

Sigmoid funksiyasining logistic regressiyadagi to'liq formulasi quyidagicha bo'ladi:

$$f(x) = \frac{1}{1 + e^{-(XA^T)}} \quad (20)$$

Sigmoid funksiyaning qisqaroq ko'rinishi quyidagicha ifodalaniladi:

$$f(Z) = \frac{1}{1 + e^{-Z}} \quad (21)$$

21-formulada $Z=XA^T$ deb hisoblanadi.

XULOSA

Maqolada maktab o'quvchilarni 10 yil davomida o'qigan fanlari va ulardan olgan baholarining ta'lim yo'nalishiga ta'sirlari o'rganilgan holda, trening ma'lumotlar to'plami hosil qilindi. Ushbu trening ma'lumotlar to'plamidan ko'p o'zgaruvchili chiziqli regressiya

funktsiyasini kiruvchi qiymatlari sifatida foydalanildi va tahlil qilindi. Tahlillar shuni ko'rsatdiki, taklif qilingan mashinaviy o'qitish modelini qurishda tanlangan qiymat funksiyasi va gradiyent tushush jarayonlari berilgan qiymatlarga eng yaqin qiymatlarni prognoz qilish imkoniyatlarini berdi. Ko'p

o'zgaruvchili chiziqli regressiyaning formulasi berilgan qiymatlarga nisbatan to'g'ri tashkil qilinganligi isbotlandi. Prognoz qilgan qiymatni ehtimolini aniqlashda sigmoid funksiyasidan foydalaniladi. Model yordamida tuzilgan algoritm va Python dasturlash tili yordamida yaratilgan dasturiy vosita

10 yil davomida matematika va fizika fanlaridan olingan baholar, o'quvchining faoligi va yutuqlari asosida maktab bitiruvchilarini kelajakda qanday ta'lim yo'nalishida ta'lim olsa, yuqori yutuqlarga erishishi mumkinligini aniqlab berdi.

FOYDALANILGAN ADABIYOTLAR

1. Zayniddinov H., Nurmurodov J., Qobilov S. Application of Machine Learning Methods for Signal Processing in Piecewise-Polynomial Bases // Proceedings - 9th IEEE International Conference on Information Technology and Nanotechnology, ITNT 2023, 2023
2. Zayniddinov H., Singh M., Tiwary U.S., Singh D. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2023, 13741 LNCS, страницы v–vi
3. Zayniddinov H., Mallayev O. Parallel Algorithm for Calculating the Learning Processes of an Artificial Neural Network // AIP Conference Proceedings this link is disabled, 2022, 2647, 050006 <https://doi.org/10.1063/5.0104178>
4. Mallayev O.U. Parallellik paradigmasi asosida tibbiy tasvirlarga raqamli ishlov berishning parallel algoritmi // Innovation in technology and science education conference. Volume 2 Issue 11. ISSN 2181-371X
5. Zayniddinov H., Mallayev O., Kuchkarov M. Parallel algorithm for modeling temperature fields using the splines method // 2021 IEEE International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021 - Proceedings, 2021, 9422645
6. Zayniddinov H., Mallayev O., Nurmurodov J. Parallel Algorithm for Constructing a Cubic Spline on Multi-Core Processors in a Cluster // 14th IEEE International Conference on Application of Information and Communication Technologies, AICT 2020 - Proceedings, 2020, 9368680
7. Zayniddinov H.N., Mallaev O.U., Anvarjonov B.B. A parallel algorithm for finding the human face in the image // IOP Conference Series: Materials Science and Engineering, 2020, 862(5), 052004
8. Волчек Ю.А., Шишко О.Н., Спиридонова О.С., Мохорт Т.В. Положение модели искусственной нейронной сети в медицинских экспертных системах // Juvenis scientia. – 2017. – №. 9.
9. Zayniddinov, H., Ibragimov, S., Tojiboyev, G., Nurmurodov, J. Efficiency of Parallelization of Haar Fast Transform Algorithm in Dual-Core Digital Signal Processors // Proceedings of the 8th International Conference on Computer and Communication Engineering, ICCCE 2021, 2021, страницы 7–12, 9467190
10. H. Zayniddinov, B. Rakhimov, G. Khalikova, A. Saidov Review and analysis of computer vision algorithms, AIP Conference Proceedings this link is disabled, 2023, 2789, 050022

