



MASHINALI O'QITISH ALGORITMLARI ORQALI TURLI SOHA OID MA'LUMOT MANBALARI INTELLEKTUAL TAHLIL QILISH

Xusenov Murodjon Zoxirovich

Buxoro davlat universiteti, e-mail: m.z.xusenov@buxdu.uz

ORCID: 0000-0002-1533-3102

Annotatsiya: Mazkur maqola, Hozirgi kunda sun'iy intellekt algoritmlari va dasturlarini ishlab chiqishda an'anaviy statistik usullardan foydalanilmoqda. Berilgan to'plamlar qiymatlari asosida regressiya va chiqizli interpolyatsiya usullaridan foydalanib kelingan. Quyida mavjud statistik usullar keltirilgan bo'lib, bular hozirgi kunda SI algoritmlari va dasturlarini ishlab chiqishda keng foydalanib kelinmoqdaligi haqida.

Kalit so'zlat: Ma'lumotlar modeli, iyerarxik modellar, relyatsion modeli, tarmoq model, tasniflash, klasterlash, sklear-learning, Tkinter, Numpy, pandas

Asosiy qism: Hozirgi vaqtda ko'plab sohalar har xil turdagi juda katta ma'lumotlar to'plamlari bilan shug'ullanadi. Barcha ma'lumotlarni qo'lda qayta ishlash ko'p vaqt talab qilishi va hatto uzoq muddatda qo'shimcha qiymat bermasligi mumkin. Mashinali o'qitish asosiy g'oyasi shundaki, kirish va chiqishlarning har qanday kombinatsiyasi o'rtasida matematik bog'liqlik mavjud. Mashinani o'rganish modeli bu munosabatlar haqida oldindan ma'lumotga ega emas, lekin agar yetarli ma'lumotlar to'plami taqdim etilsa, uni yaratishi mumkin. Bu shuni anglatadiki, har bir mashinani o'rganish algoritmi o'zgartirilishi mumkin bo'lgan matematik funktsiya atrofida qurilgan. Mashinali o'qitish algoritmlari bashorat qilingan qiymat yoki sinfga ta'sir qiluvchi har bir mustaqil o'zgaruvchi uchun eng yaxshi qiymatni topish uchun ma'lumotlarga o'rgatadi. Hozirgi vaqtda ko'plab sohalar har xil turdagi juda katta ma'lumotlar to'plamlari bilan shug'ullanadi. Barcha ma'lumotlarni qo'lda qayta ishlash ko'p vaqt talab qilishi va hatto uzoq muddatda qo'shimcha qiymat bermasligi mumkin.

Mashinali o'qitish ma'lumotlar tahlili modelni yaratishni avtomatlashtiradigan usullardan biri. Bu mashinalar tajriba orqali o'rganishi va moslasha olishi kerak degan fikrga asoslangan sun'iy intellekt sohasi. Bu kompyuter tomonidan to'plangan statistik ma'lumotlarga asoslanib bashorat qiladigan hisoblash statistikasi bilan chambarchas bog'liq. Bu ba'zan ma'lumotlarni tanlab olish bilan aralashib ketadi, lekin u ko'proq razvedka ma'lumotlarini tahlil qilishga qaratilgan bo'lsa, mashinani



o'rganish asosan bashorat qilish uchun ishlatiladigan murakkab algoritmlarni o'z ichiga oladi.

Mashinani o'rganish usullari yordamida hal qilinadigan vazifalarning asosiy qismi ikki xil turga tegishli: o'qituvchi bilan o'rganish (nazorat ostida o'rganish) yoki usiz (nazoratsiz o'rganish). Biroq, bu o'qituvchi kompyuter ustida turib, dasturdagi har bir harakatni boshqaradigan dasturchining o'zi bo'lishi shart emas. Mashinani o'rganish nuqtai nazaridan "o'qituvchi" — bu ma'lumotni qayta ishlash jarayoniga shaxsning aralashuvi. O'rganishning ikkala turida ham mashina tahlil qilish va qonuniyatlarni topishi kerak bo'lgan dastlabki ma'lumotlar bilan ta'minlangan. Mashinali o'qitishning ko'plab usullari mavjud: Shulardan ikkitasi ko'rib o'tamiz

1. **Nazorat ostida o'qitish**- agar natijalarning kirishlarga funktsional bog'liqligini topish va kirishda ob'ekt tavsifini qabul qiladigan va chiqishda javob beradigan algoritmi qurish kerak bo'lganda. Sifat funktsionalligi, qoida tariqasida, namunadagi barcha ob'ektlar uchun algoritmi javoblarning o'rtacha xatosi orqali aniqlanadi. Nazorat ostidagi o'rganish tasniflash, regressiya, tartiblash va bashorat qilish muammolarini o'z ichiga oladi.

2. **Nazoratsiz o'rganish** — ma'lumotlarini tekshiradi, ularning aksariyati belgilanmagan va tuzilmagan va mavjud bo'lgan barcha tegishli ma'lumotlardan foydalangan holda naqsh va qonuniyatlarni aniqlay aniqlashni boshlaydi.

Tasniflashni boshlash uchun bir necha qadamlar:

Muammoni tushunish: Tasniflashni boshlashdan oldin, siz hal qilmoqchi bo'lgan muammoni tushunish muhimdir. Siz bashorat qilmoqchi bo'lgan sinf belgilari qanday? Kirish ma'lumotlari va sinf belgilari o'rtasida qanday bog'liqlik bor? Ma'lumotlarni tayyorlash: Muammoni yaxshi tushunganingizdan so'ng, keyingi qadam ma'lumotlaringizni tayyorlashdir. Bunga ma'lumotlarni yig'ish va qayta ishlash, shuningdek, ularni o'qitish, tekshirish va test to'plamlariga bo'lish kiradi.

Modelni tanlash: Tasniflash uchun ishlatilishi mumkin bo'lgan juda ko'p turli xil modellar mavjud, jumladan qaror daraxtlari, tasodifiy o'rmonlar, keng yaqin qo'shnilar va qo'llab-quvvatlovchi vektor mashinalari. Ma'lumotlaringiz hajmi va murakkabligi hamda mavjud hisoblash resurslarini hisobga olgan holda muammoingizga mos modelni tanlash muhimdir.

Modelni o'rgatish: Modelni tanlaganingizdan so'ng, keyingi qadam uni o'quv ma'lumotlariga o'rgatishdir. Bu o'quv ma'lumotlari uchun taxmin qilingan sinf yorliqlari va haqiqiy sinf yorliqlari o'rtasidagi xatolikni minimallashtirish uchun model parametrlarini sozlashni o'z ichiga oladi.



Modelni baholash: Modelni o'rgatgandan so'ng, uning ishlashini tekshirish to'plamida baholash muhimdir. Bu sizga modelning yangi, ko'rilmagan ma'lumotlarda qanchalik yaxshi ishlashi haqida yaxshi fikr beradi. Modelni sozlash: Agar modelning ishlashi qoniqarli bo'lmasa, siz parametrlarni sozlash yoki boshqa modelni sinab ko'rish orqali uni nozik sozlashingiz mumkin.

Modelni joylashtirish: Nihoyat, modelning ishlashidan qoniqsangiz, uni yangi ma'lumotlar bo'yicha bashorat qilish uchun ishlatishingiz mumkin. Tasniflash algoritmlarini 3ta asosiy guruhi mavjud:

1.Chastotsli

- Naive Bayes
- Qarorlardaraxti

2.Koverians matritsa

- Chiziqli diskiriminant tahlil
- Logistik regressiya

3.O'xshashlik funksiyalari

- Keng yaqin qo'shnilar.

Naive Bayes algoritmi- Bayes teoremasiga asoslanadi va va tasniflash masalalarini hal qilish uchun ishlatiladigan nazorat ostida o'rganish algoritmi. Bayes teoremasining matematik formulasi quyidagicha berilgan:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} [1]$$

bu yerda A va B hodisalar va $P(B) \neq 0$. Biz B hodisasi to'g'ri ekanligini hisobga olib, A hodisasining ehtimolini topishga harakat qilamiz.

$P(A)$ - A hodisaning (apriori) oldingi ehtimollik, ya'ni dalil ko'rinmasdan oldin sodir bo'lgan voqea ehtimoli. $P(A|B)$ — B hodisaning (posteriori ehtimoli), ya'ni dalil ko'rilgandan keyin sodir bo'lish ehtimoli. Teorema oldin sodir bo'lgan hodisa ehtimoligino hisobga olgan holda voqea sodir bo'lish ehtimolini topadi.

Bayes teoremasi sinf o'zgaruvchisi berilgan quyidagi munosabatini bildiradi:

$$P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n | y)}{P(x_1, \dots, x_n)}$$

Qaror daraxtlari- tasniflash va regressiya muammolarini hal qilishga imkon beruvchi eng kuchli ma'lumotlar qazib olish va bashoratli tahlil vositalaridan biridir. Qaror daraxtlarining asosiy vazifasi statistika, ma'lumotlarni tahlil qilish va mashinani o'rganishda qo'llaniladigan boshqaruv qarorlarini qabul qilish jarayonlarini qo'llab-quvvatlashdir. Qarorlar daraxti berilgan jadval ko'rinishidagi ma'lumotlarda har bir ustun bo'ylab turli xil qoidalarni sinab ko'radi va eng



maqbulini oladi. Eng maqbulini esa o'z-o'zidan qoidalar asosida yaratilgan bo'linma to'plamlar orasidagi ma'lumotlar entropiyasini baholash orqali tanlab olinadi. Ma'lumotlar to'plami entropiyasi bu- berilgan to'plamdagi ma'lumotlarning o'rtacha barqarorligini aniqlovchi o'lchov. Ya'ni berilgan to'plamdagi ma'lumotlar turlari qay darajada ko'p? degan savol javob beradi va quyidagi formula orqali hisoblanadi .

$$\text{Entropiya} = \sum_i^n P = p_i p_i$$

P_i - to'plamdagi i -klassning ehtimoli ya'ni i -klassdagi elementlar sonining to'plamdagi barcha elementlar soniga bo'linmasi.

Tasniflash — ob'ektlarni oldindan ma'lum bo'lgan sinflardan biriga belgilash. Maqsadli o'zgaruvchi diskret qiymatlarga ega bo'lishi kerak.

Regressiya (raqamli bashorat)— berilgan kirish vektori uchun mustaqil o'zgaruvchining raqamli qiymatini bashorat qilish.

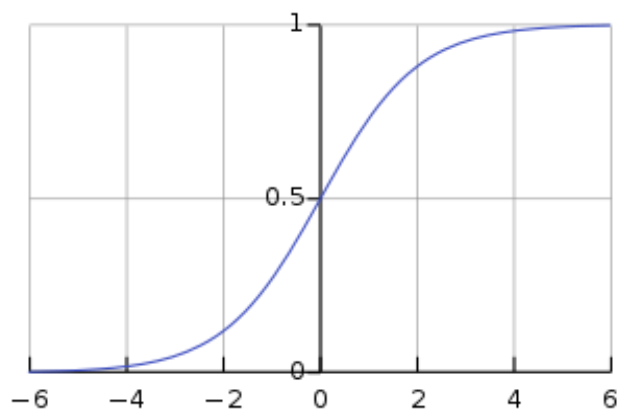
Ob'ektlar tavsifi— qarorlar daraxtidagi qoidalar to'plami ob'ektlarni ixcham tasvirlash imkonini beradi. Shuning uchun, ob'ektlarni tavsiflovchi murakkab tuzilmalar o'rniga siz qaror daraxtlarini saqlashingiz mumkin.

Chiziqli diskriminant tahlil- ikki yoki undan ortiq sinflarga tegishli naqshlarni ajratish uchun ishlatiladi. [2] Nazorat qilinadigan tasniflash muammolari uchun ishlatiladigan o'lchamlarni kamaytirish usulidir. U guruhlardagi farqlarni modellashtirish, ya'ni ikki yoki undan ortiq sinflarni ajratish uchun ishlatiladi.

Logistik regressiyadan maqsad diskret (ya'ni faqat ma'lum turdagi) nishonlarni bashorat qiladi. [3]. Tasniflash muammolari doirasida bizda kirish o'zgaruvchilari (X) va chiqish o'zgaruvchisidan (y) iborat etiketli o'quv ma'lumotlar to'plami mavjud. Logistik regressiya algoritmi X va y o'rtasidagi munosabatlarni tavsiflash uchun eng mos logistik funksiyani topishga yordam beradi.

Logistik regressiya nima ekanligini va u qanday ishlashini tushunish uchun sigmasimon funksiya va tabiiy logarifm funksiyasini tushunishingiz kerak bo'ladi . Logistik funksiya tenglamasi yoki logistik egri chiziq quyidagi tenglama bilan aniqlangan umumiy "S" shaklidagi egri chiziqdir. Logistik egri chiziq

sigmasimon egri deb ham ataladi. $S(X) = \frac{1}{1 + e^{-k(x-x_0)}}$



Logistik regressiya uchun AQSHning diabet, oshqozon-ichak va buyrak kasalligilari milliy instituti ma'lumotlariga asoslangan ma'lumotlar jadvalida aholining 21 yoshidan katta bo'lgan ayollardagi turli analizlar, xususan diabet kasalligi bor yo'qligi keltirilgan jadvaldan foydalanamiz. Ma'lumotlar to'plami bir nechta tibbiy bashorat qiluvchi (mustaqil) o'zgaruvchilardan va bitta maqsadli (qaram) o'zgaruvchidan, Natijadan iborat. Mustaqil o'zgaruvchilarga bemorning homiladorlik soni, ularning BMI, insulin darajasi, yoshi va boshqalar kiradi.

Triceps teri burmasining qalinligi (mm)	2 soatlik sarum insulini (mu U/ml)	Tana massasi indeksi (vazn kg/(bo'yi m)^2)	Qandli diabetning naslchilik funksiyasi	Yoshi (yillar)	Sinf o'zg 1) 768 ta qolganla
0 99	0 846	0 67.1	0,08 2.42	21 81	0
0	0	27.1	1.441	57	0
23	846	30.1	0,398	59	1
19	175	25.8	0,587	51	1
0	0	30	0,484	32	1

Logistik regressiya muammosi o'rganish uchun kerakli kutubxonalarni o'rnatamiz: import numpy as np import pandas as pd import matplotlib.pyplot as plt



from matplotlib.pyplot import figure import math //Endi yuqoridagi bazani pandas kutubxonasi yordamida yuklaymiz: df=pd.read_csv('diabetes.csv') //endi ma'lumotlar bazasi bilan tanishib chiqamiz df=head() //endi ma'lumotlar to'plamining texnik jihatlari bilan tanishsak ya'ni qanday formatda , qatorlar soni va h/k bilan tanishsak df.info() Ma'lumotlar to'plami 768 qatordan iborat bo'lib turli xil formatlar ega (int64 va flot64) . Ushbu ma'lumotlar to'plami matematik amallar, xususan logistik regressiyaga moslash uchun ustunlar formati bir xil qilish kerak: df=np.array(df,type=float)

K-Eng yaqin qo'shnilar (K-NN)- algoritmi shunga o'xshash narsalar yaqin joyda mavjudligini taxmin qiladi. Boshqacha qilib aytganda, o'xshash narsalar bir-biriga yaqin. Algoritmning afzalligini oddiy misolda tushuntiramiz. Aytaylik, ikkita ma'lumotlar to'plami ya'ni A va B to'plam mavjud va bizda yangi ma'lumotlar kelib qo'shildi, bu ma'lumotlar ushbu ma'lumotlar to'plamidan qaysi birida tegishli ekanligini hal qilish kerak. Ushbu turdagi muammolarni hal qilish uchun bizga K-NN algoritmi kerak. K-NN yordamida biz ma'lum bir ma'lumotlar to'plamining toifasi yoki sinfini osongina aniqlashimiz mumkin.

Foydalanilgan adabiyotlar:

1. Ye N. Data Mining: Theories, Algorithms, and Examples // Data Mining: Theories, Algorithms, and Examples. 2013. 31 p.
2. Devi M.N.M, V.S. Pattern Recognition An Algorithmic Approach. 162 p.
3. S.Kamolov, Sun'iy intellekt asoslari. mashinaviy o'qitish. 2022. 46 p.
4. M.Z.Xusenov, L.O.Sharipova, Oliy ta'lim muassasalarida masofaviy ta'limni joriy qilish, "Pedagogik mahorat" ilmiy-nazariy va metodik jurnal. 2022, № 2 B: 94-96
5. M.Z.Xusenov, L.O.Sharipova, Kimyo fanini o'qitishda VR texnologiyasini qo'llash, Pedagogik mahorat Ilmiy-nazariy va metodik jurnal maxsus son (2021-yil, dekabr) 164-166
6. M.Z.Khusenov, L.O.Sharipova, Statistical analysis of network problems and their impact on the practice of social computing in Uzbekistan, E3S Web of Conferences Volume 389, 09017 (2023) Ural Environmental Science Forum "Sustainable Development of Industrial Region-31 May 2023