**ISOQOV RO‘ZIMUROD (110-20)**

ML.1-modul(amaliy)

Mashinali o'qitishning asosiy tushunchalari kompyuterlar yordamida o'rganish, tahlil qilish va natijalarni tushunishga qaratilgan ilmiy yo'nalishdir. Bu tushunchalar boshqa sohalardan foydalanish yoki yangi model yaratishda yordam beradi. Mashinali o'qitish algoritmlari aldovlar to'plami emas, balki kompyuter uchun yopiq tushunchalar, xuddi shu qanday mashinalar yaratilish va ular bilan ishlash asosida o'rganiladi.

***Mashinali o'qitishning asosiy tushunchalari va algoritmlari*** quyidagilardan iborat:

1. **Model yaratish:** Mashinali o'qitishda birinchi navbatda, o'rganishni xohlagan ma'lumotlarni o'z ichiga olgan model yaratiladi. Bu model, ma'lumotlar bilan ishlashni va ularga aniq qarorlar qabul qilishni o'rganadi.
2. **Ma'lumot tayyorlash:** Model uchun ma'lumotlar to'plami tayyorlanishi kerak. Ma'lumotlar soni, xususiyatlari, formati va qismlash usullari ma'lumotlar tayyorlash jarayonida muhim.
3. **Algoritmlarni tanlash:** Mashinali o'qitishda, bir necha algoritm turini ishlatish mumkin. Algoritm tanloviga ko'proq vaqt sarflanadi, chunki har bir algoritmda o'ziga xos qulayliklar va cheklovlar mavjud.
4. **Modelni o'qitish:** Model uchun ma'lumotlar tayyorlanganda va algoritm tanlanib, bir necha yuritilgan algoritmlarni bajarib, model o'qitiladi. Bu, modelning aniqroq va ishonchli qarorlar qabul qilishini ta'minlaydi.
5. **Model sinovlari:** O'qitilgan model sinov jarayonidan o'tadi. Bu jarayonda, sinov ma'lumotlari modelga kiritiladi va modelning qarorlarining aniqligi va aniqroqligi sinovlanadi.
6. **Modelni yaxshilash:** Model sinovdan o'tqazilgandan so'ng, uni yaxshilash maqsadida bir nechta qo'shimcha sinovlar o'tkaziladi. Bu sinovlar o'qitilgan modelni yaxshilashda va optimallashtirishda yordam beradi.

Mashinali o'qitishning asosiy tushunchalari va algoritmlari shu tartibda amalga oshiriladi va bu jarayon odatda ko'p marta takrorlanadi, joriy natijalardan tashqari.

***Mashinali o'qitish turli*** xil turlardan iborat bo'lishi mumkin:

1. **Supervized (*mudofaa qilinadigan*) o'qitish:** Bu mashinali o'qitishning eng ko'p ishlatiladigan turi hisoblanadi. Bu usulda, o'qituvchi modellarga belgilangan ma'lumotlar beriladi va ular qarorlar qabul qilishda ishlatiladi. Bu ma'lumotlar "*to'g'ri javob*" yoki " *noto'g'ri javob*" holatlarida beriladi.
2. **Unsupervised (*mudofaa qilinmaydigan*) o'qitish:** Bu mashinali o'qitishning boshqa turi hisoblanadi. Bu usulda, o'qituvchilar modellarni "otash" qilish orqali o'rganishga yordam beradi. Bu usulda ma'lumotlar "*to'g'ri javob*" yoki "*noto'g'ri javob*" holatlarida berilmaydi.
3. **Reinforcement (*takrorlovchi*) o'qitish:** Bu mashinali o'qitishning boshqa turi hisoblanadi. Bu usulda, modellarga belgilangan vazifalarni bajarish uchun "*takrorlovchi o'qitish*" ishlatiladi.

***Mashinali o'qitish jarayonining umumiy qadamlari*** quyidagilardan iborat:

1. **Ma'lumotlarni to'plash:** Mashinali o'qitish uchun, muhim ma'lumotlar to'planadi. Bu ma'lumotlar tashqi ma'lumotlar bazalari yoki so'nggi yillarning malakali operatorlar tomonidan yig'ilgan ma'lumotlar bo'lishi mumkin.
2. **Ma'lumotlarni tayyorlash:** Ma'lumotlarni o'qitish uchun to'plaganida, ularni tayyorlash kerak. Bu, ma'lumotlar ko'rinishi, cheklovlar, format va qismlash usullari bo'lishi mumkin.
3. **Modelni yaratish:** Ma'lumotlar to'plagan va tayyorlangandan so'ng, model yaratiladi. Modelning qanday ishlashini, ko'rsatuvchi bir necha xususiyatlar va qarorlar qabul qilishni o'rganish uchun, o'qitish uchun ma'lumotlar beriladi.
4. **O'qitish:** Ma'lumotlar modelga kiritiladi va model qarorlar qabul qilish uchun o'rganiladi. O'qitilgan modelning xususiyatlari va jarayonini boshqarish uchun belgilangan "*o'qitish*" algoritmlari mavjud.
5. **Modelni testlash:** Model o'qitilgandan so'ng, uning aniqligi va nisbatan tayyorligini tekshirish uchun ma'lumotlarni testlash va tasdiqlash kerak. Bu testlash ko'p qatlamli bo'lishi mumkin va bu testlash jarayoni "*testlash to'g'risida*" ma'lumotlar bazasida saqlanishi mumkin.
6. **Modelni ishlatish:** Modelning maqsadga muvofiqligi tasdiqlanganidan so'ng, u amaliyotda ishlatilishi mumkin. Bu o'nlab ishlar uchun ishlatilishi mumkin, masalan: ma'lumotlar ustidan taqdimotlar yaratish, ma'lumotlar ustida tahlil va boshqa maqsadlar.
7. **Modelni yangilash:** Ma'lumotlar va ulardan foydalanish usullari ko'payib, modellarning aniqligi va ko'rsatkichlarida o'zgarishlar bo'lishi mumkin. Modelni o'zgartirish va yangilash bu jarayonning bir qismidir, chunki modellarning ko'rsatkichlarini yaxshilash va yangi ma'lumotlarga mos keluvchi xususiyatlar qo'shish mumkin.

Bu jarayonning har bir qismi mavjud modellar va ma'lumotlarga qarab o'zgartirilishi mumkin. Biroq, asosiy adabiyotlarni va qadamlarni tushunishning yaxshi bir qismi, qisqa vaqt ichida barcha qadamlarni bajarish uchun yuqori darajada avtomatlashtirilgan mashinali o'qitish (AutoML) xizmatlari mavjudligidir.

***Mashinali o‘qitishda instrumental vositalardan foydalanish, Google colab dasturiy muhiti bilan ishlash.***

Google Colab, Google tomonidan taqdim etilgan, bulut tabaqalashtirilgan dasturiy muhitidir. Bu muhit, Python dasturlash tili yordamida mashinali o'qitish, tahlil va boshqa matematik hisob-kitoblari yozish uchun xizmat qiladi. Colab, shaxsiy kompyuterda Python va kerakli kutubxonalarni o'rnatish, yoki kerakli konfiguratsiyalarni o'rnatish va sozlashni talab qilmaydi.

Colab-da Python skriptlarini yozish vaqti kelganida, kerakli kutubxonalarni o'rnatish, yuklash va sozlash jarayonini boshqarish uchun, qulay interfeysga ega bo'lgan klavishlar mavjud. Bunda jumladan, Google Colab, ana shu interfeys orqali foydalanuvchilar uchun tayyorlangan maksimal darajada samarali vosita hisoblanadi.

Mashinali o'qitishda Google Colab-ni ishlatish oson va samarali. Google Colab muhitida foydalanuvchi, Kerakli kutubxonalarni, dasturlash tilini va talablarni o'rnatishni va Python skriptlarini yozishni o'rganishi mumkin. Google Colab muhitida o'zgaruvchilar, matritsalar va statistik ma'lumotlar ishlash uchun ishlatilgan kutubxonalarga, shuningdek, skriptlar ishga tushirish uchun kerakli buyruqlar, funksiyalar va tizimlar mavjud.

Mashinali o'qitishda Google Colab, kerakli kutubxonalarni yuklash, ma'lumotlar bazasiga ulashish va ma'lumotlarni tahlil qilish, grafik yaratish, tizimni o'zgartirish va yangilash, xususiyatlar tanlash va modellarni sinovlash va baholash uchun ko'p xil qulayliklar yaratadi. Google Colab foydalanuvchilariga o'zlarining modellari uchun, yuklab olishga talab qilinadigan to'plamlar tayyorlash va ularni foydalanishni osonlashtiradi. Buning yanada ko'proq vaqt va xabar to'plamlarini to'playdigan jihatlaridan foydalanib, mashinali o'qitishni avtomatlashtirishga qodir bo'ladi.

**Google Colab-ni ishlatishning afzalliklari va kamchiliklari mavjud.**

***Afzalliklari:***

* Ilova bulutda joylashtirilganligi uchun, shaxsiy kompyuterda Python va boshqa muhitlarni o'rnatish va sozlash talab qilmaydi.
* Google Colab-ni ishlatish bepuldir.
* Google Colab Jupyter Notebookning xususiyatlariga ega va interaktivlikni ko'proq ko'rsatadi.
* Google Colab Googlening xususiy tizimlarini, masalan, Google Drive, Google Sheets, Google BigQuery va boshqalarini ishlatishga imkon beradi.
* Google Colab bir nechta foydalanuvchilar tomonidan birgalikda ishlatilishi mumkinligi uchun birgalikda ishga tushirish va baham ko'rish funksiyalarini o'z ichiga olgan.
* Google Colab PyTorch, TensorFlow, Keras va boshqa mashhur kutubxonalarni ko'rsatadi.
* Google Colab GPU va TPU yoritgichlari bilan ishlashni osonlashtiradi.

***Kamchiliklari:***

* Google Colab onlayn muhitda ishlash uchun internetga ulanish talab qiladi.
* Google Colab bir nechta foydalanuvchilar tomonidan ishlatilganda, ishlab chiqarish jarayonida cheklovlar paydo bo'lishi mumkin.
* Google Colab sinov jadvallari va modellar kichik ma'lumotlar uchun mos keladi. Katta modellar va ko'p xil data setlari uchun, lokal kompyuterning xususiyatlaridan ko'ra tez ishlamaydi.
* Google Colab-da uzoq vaqt ishlashi mumkin bo'lgan jarayonlar uchun, Chrome brauzeri orqali ulanishga tegishli qo'shimchalar o'rnatilishi kerak.
* Google Colab-da ishlatiladigan resurslar, boshqa foydalanuvchilar bilan baham ko'rish tufayli cheklanishi mumkin.

Barcha kamchiliklarga qaramay, Google Colab-ni ishlatish, kerakli buyruq va funksiyalar yozish va mashinali o'qitish va tahlilning ko'plab qismlarini o'rganish uchun yaxshi imkoniyat yaratadi.

***Bir o'zgaruvchili va ko'p o'zgaruvchili chiziqli regressiya*** masalalari mashinali o'qitishda katta ahamiyatga ega bo'lgan masalalardan hisoblanadi. Bu masalalar ko'p xil sohalar uchun foydalaniladi, masalan, ma'lumotlar tahlili, xavfsizlik, sotsial so'nggi voqealar va boshqalar. Bu regressiya turlarini dasturlashning turli usullari bilan hal qilish mumkin.

***Bir o'zgaruvchili chiziqli regressiya*** masalasi, yagona o'zgaruvchini mavjud ma'lumotlar to'plamiga mos kelgan yo'ldan boshqa o'zgaruvchilarga nisbatan boshqarishga harakat qiladi. Misol uchun, bir shaxsning bo'yi va og'irligi mavjud bo'lsa, uning yoshini bashorat qilish uchun uning bo'yi va og'irligi parametrlariga mos keladigan model yaratish mumkin.

***Ko'p o'zgaruvchili chiziqli regressiya*** masalalari esa, bir nechta o'zgaruvchilarni mavjud ma'lumotlar to'plamiga mos kelgan yo'ldan boshqa o'zgaruvchilarga nisbatan boshqarishga harakat qiladi. Misol uchun, bir shaxsning xavfsizlik riskini bashorat etish uchun uning yoshi, jinsi, tajribasi va boshqa ko'plab o'zgaruvchilarni o'z ichiga olgan model yaratish mumkin.

Bu regressiya turlarini dasturlash uchun Python va *NumPy*, *Pandas*, *Matplotlib* va *Sklearn* kabi kutubxonalarni ishlatish mumkin. Ma'lumotlar to'plamlarini yuklash, ko'rib chiqish va tahlil qilish uchun *Pandas* kutubxonasi foydalaniladi. *Model* yaratish va sinov jarayonini amalga oshirish uchun *Sklearn* kutubxonasi foydalaniladi. *Matplotlib* kutubxonasi grafiklarni yaratish uchun foydalaniladi.

Chiziqli regressiya masalalari ko'p vaqtda *Gradient Descent* va *Stochastic Gradient Descent* kabi o'zgaruvchi o'rganish algoritmlari yordamida yechiladi. Yechimni baham ko'rish va modellarni sinash uchun **MSE** (*Mean Squared Error*), **R-Squared** va **Adjusted R-Squared** kabi metrikalar foydalaniladi.

Quyidagi regressiya turlari keng tarqalgan regressiya turlaridir:

* **Bir o'zgaruvchili chiziqli regressiya:** Bu regressiya turi, yagona o'zgaruvchini mavjud ma'lumotlar to'plamiga mos kelgan yo'ldan boshqa o'zgaruvchilarga nisbatan boshqarishga harakat qiladi. Misol uchun, bir shaxsning bo'yi va og'irligi mavjud bo'lsa, uning yoshini bashorat etish uchun uning bo'yi va og'irligi parametrlariga mos keladigan model yaratish mumkin.
* **Ko'p o'zgaruvchili chiziqli regressiya:** Bu regressiya turi, bir nechta o'zgaruvchilarni mavjud ma'lumotlar to'plamiga mos kelgan yo'ldan boshqa o'zgaruvchilarga nisbatan boshqarishga harakat qiladi. Misol uchun, bir shaxsning xavfsizlik riskini bashorat etish uchun uning yoshi, jinsi, tajribasi va boshqa ko'plab o'zgaruvchilarni o'z ichiga olgan model yaratish mumkin.
* **Polinomial regressiya:** Bu regressiya turi, to'g'ridan iborat bo'lmagan har xil xossalarni (masalan, yosh, bo'yi, og'irlik va boshqalar) parametrlar sifatida ishlatadi va ularga yuqori darajadagi o'lchamli o'zgaruvchilar qo'shadi. Polinomial regressiya, mos keladigan funksiyani chizishga imkon beradi va ma'lumotlar to'plamiga mos keladigan modelni yaratishga yordam beradi.
* **Regularizatsiyali regressiya:** Bu regressiya turi, o'zgaruvchilar miqdorini va yuqori darajadagi o'lchamli o'zgaruvchilarni cheklovchi regularizatsiya (L1 va L2) ishlatish yordamida o'zgaruvchilarning kuchini kamaytiradi. Bu regressiya turi, overfittingni kamaytirishga yordam beradi va modelning umumiy ishonchini oshiradi.
* **Logarifmik regressiya:** Bu regressiya turi, mazkur funksiyani xossalarni mavjud ma'lumotlarga mos kelgan ko'paytiruvchi o'zgaruvchilarga nisbatan boshqarishga harakat qiladi. Bu regressiya turi, kategoriya ko'paytiruvchilari va ularning qiymatlari mavjud bo'lsa, uni aniqlash uchun ishlatiladi.
* **Ridge regressiya:** Bu regressiya turi, L2 normasidan foydalanadi va yuqori darajadagi o'lchamli o'zgaruvchilarni cheklovchi regularizatsiya ishlatadi. Bu regressiya turi, overfittingni kamaytirish uchun ishlatiladi va o'zgaruvchilarning qatorining kuchini cheklaydi.
* **Lasso regressiya:** Bu regressiya turi, L1 normasidan foydalanadi va katta sonli o'zgaruvchilarning miqdorini kamaytiradi. Bu regressiya turi, o'zgaruvchilarning turini cheklovchi regularizatsiyalar ishlatadi va shunchaki o'zgaruvchilar to'plamining katta qismini yopadi.

Bularning har biri, ma'lumotlardagi yo'qolishlarni bashorat etish uchun mos keladigan modelni yaratish uchun foydalaniladi. O'zgaruvchilar turlari va cheklovchi jazbalar miqdori, modellarni baholash uchun ko'p qo'llaniladi va modellarni to'g'ri sozlash va ishga tushirish uchun ma'lumotlarni tasniflash, o'rganish va ishlab chiqish jarayonlarini tahlil qilishda foydalaniladi.

***Logistik regressiya tushunchasi va ularning mashinali o‘qitishda qo‘llanilishi.***

**Logistik regressiya** klassifikatsiya uchun ma'lumotlar ustidan qarama-qarshi keluvchi natijalarni bashorat etish uchun ishlatiladigan bir regressiya turi hisoblanadi. Bu regressiya turi, odatda ikki klassga ajratilgan natijalarga ega bo'lgan ma'lumotlarda ishlatiladi: *pozitiv* (1) va *negativ* (0).

Misol uchun, sizning ma'lumotlar bazangizda mijozlarning bir xil mahsulotni sotib olishi yoki sotmaganligi bo'yicha natijalar mavjud bo'lishi mumkin. Logistik regressiya, buyurtmani qabul qilish yoki rad etishni bashorat etishda foydalaniladi. Bu regressiya turi, ikki klassga bo'linadi: bitta klass (binary) va ko'p klassli (multiclass). Bitta klass regressiyasi faqat ikki klassga (1 va 0) ajratilgan natijalarga ega bo'lgan ma'lumotlar uchun ishlatiladi, ya'ni "*Ha*" va "*Yo'q*" yoki "*To'g'ri*" va "*Noto'g'ri*" kabi.

Logistik regressiya, yuqori darajadagi funksiyalarni ishlatadi va natijani bashorat qilishda foydalaniladigan modelning xossalari boshqa regressiya turlari bilan o'xshash. Bu turlarning har birida ko'paytirish koeffitsiyentlarini aniqlash uchun yuqori darajadagi funksiyalar ishlatiladi. Bu regressiya turi, ma'lumotlardagi klasslarning miqdorini aniqlash uchun qo'llaniladi.

Mashinali o'qitishda, logistik regressiya, klassifikatsiya uchun ma'lumotlar ustida yuritiladigan amaliyotlar uchun juda foydali bo'lishi mumkin. Bu regressiya turi, ma'lumotlarni to'g'ri sozlash va aniqlash uchun yuqori natijalarni yoki natijalarning ko'rinishlarini aniqlashda ishlatiladi. Mashinali o'qitishda, logistik regressiya o'zgaruvchilarning cheklovini aniqlash uchun ishlatiladi va bitta klass regressiyasining yordamida yuqori darajadagi funksiyalarning hammasini aniqlash uchun ishlatiladi.