

Machine Learning

Davronbek Malikov

1 3-dars: Supervised ML

Qo'lingizdagi material *Supervised Machine learning* mavzusini mustahkamlash uchun qo'shimcha ma'lumotlar va savol-javoblardan iborat.

Supervised ML 2 turga bo'linadi: Classification and Regression

1.1 Classification

Classification nima? Classification bu ML ning nazoratli o'rganish turi bo'lib, bunda model kiritilgan ma'lumotlar asosida ob'ektni turli toifalarga (sinflarga) ajratadi. Ya'ni, u natijani oldindan belgilangan sinflardan biriga ajratishga harakat qiladi. Classification muammosida natija doim diskret (qat'iy sinflangan) bo'ladi.

Classification muammolariga misollar:

- Elektron pochta xabarini spam yoki spam bo'lmagan xabar sifatida tasniflash
- Suratdagi hayvonni it yoki mushuk sifatida tasniflash
- Xastalik alomatlariga qarab bemorning kasalligini aniqlash (yurak kasalligi yoki yo'q)

Quyidagi modellardan foydalangan xolda Classification muammolarini yechishimiz mumkin. Classification modellari:

- Logistic Regression
- Decision Tree
- Naive Bayes
- KNN

Masalan: Quyida keltirilgan bo'ri yoki itni aniqlash modeli klassifikatsiya muammosidir, chunki natija ikkita toifa: "bo'ri" yoki "it". Quyidagi jadvalda berilgan rasmga ko'ra it yoki bo'rini aniqlash uchun to'plangan ma'lumotlar ni bir qismi bilan tanishishingiz mumkin bu fan tilida **data set** deb ataladi.

Table 1: Itni yoki bo'rini aniqlash uchun ma'lumotlar jadvali

Rasm ID	tana rangi	Dum uzunligi (sm)	Quloq shakli	Ko'z kattaligi (sm)	Natija (bo'ri yoki it)
1	Oq	30	O'tkir	2.5	Bo'ri
2	Jigarrang	25	Dumaloq	1.8	It
3	Kulrang	35	O'tkir	2.2	Bo'ri
4	Qora	28	Dumaloq	1.9	It
5	Oq	32	O'tkir	2.4	Bo'ri

Natija: Yuqoridagi jadvalga ko'ra boshqa features(xususiyatlar) berilganda rasmni bo'ri yoki it ekanligini aniqlay olamiz

1.2 Regression

Regression nima? Regression bu ham mashinali o'rganishning nazoratli o'rganish turi bo'lib, lekin bunda natija uzluksiz qiymat shaklida bo'ladi. Ya'ni, regression modelining vazifasi kiritilgan ma'lumotlar asosida uzluksiz miqdoriy natijani bashorat qilishdir.

Regression muammolariga misollar:

- Uy narxini bashorat qilish (uy maydoni, xonalari soni, joylashuvi kabi ma'lumotlar asosida)
- Aholi sonini prognoz qilish(2050 yilda O'zbekiston aholisi qancha bo'ladi)
- Havo haroratini oldindan aytish

Quyidagi modellardan foydalangan xolda Regression muammolarini yechishimiz mumkin. Regressiya modellari:

- Linear Regression
- Decision Tree(Regressionda xam ishlatishimiz mumkin)
- Ridge Regression

Masalan: Quyida keltirilgan ob-havoni bashorat qilish jadvali regressiya muammosidir, chunki natija uzluksiz qiymat, ya'ni havo harorati ($^{\circ}\text{C}$).

Table 2: Ob-havoni bashorat qilish uchun ma'lumotlar jadvali

Namlilik (%)	Shamol tezligi (km/soat)	Bulut qoplami (%)	Bosim (hPa)	Harorat ($^{\circ}\text{C}$)
85	10	75	1012	15
65	15	40	1015	22
70	8	20	1018	25
90	12	80	1008	12
60	20	30	1020	28

Natija: Yuqoridagi jadvalga ko'ra boshqa features(xususiyatlar) berilganda Keyingi kun uchun ob xavoni predict qila olamiz

Quyidagi jadvalda Classification va Regressionlarni farqini bilishimiz uchun jadval berilgan

Table 3: Classification va Regression taqqoslash jadvali

Xususiyat	Classification	Regression
Ma'lumot turi	Diskret sinflar (kategoriyalar)	Uzluksiz qiymatlar
Maqsad	Ob'ektlarni belgilangan sinflarga ajratish	Miqdoriy qiymatlarni bashorat qilish
Natija turi	Toifalar (masalan, "bo'ri" yoki "it")	Uzluksiz qiymatlar (masalan, havo harorati: 25°C)
Misol	Elektron xabarni spam yoki spam emas deb tasniflash	Uy narxini (so'mda) bashorat qilish
Modellar	Logistik regression, Naive Bayes, Decision Tree	Linear Regression (Regression Trees)
Foydalanish sohasi	Tibbiy tashxis, suratlar tasniflash, hujjatlarni toifalash	Narxlarni prognoz qilish, havo haroratini bashorat qilish
Natijaga misol	Ob'ekt tasnifi: It yoki Mushuk	Miqdoriy qiymat: 500 000 so'mlik uy narxi

2 MI model yaratishimiz uchun kerak bo'ladigan vositalar

MI topshiriqlarini bajara olishimiz uchun yuqoridagi o'rgan bilimlarimizga qo'shimcha vositalar kerak bo'ladi. Masalan Python bu dasturlash tillaridan biri bo'lib MI uchun yaqin do'st kodlarni o'rnatishimiz va natijalarni amalda ko'ra olishimiz uchun kerak bo'ladi. Undan tashqari boshqa vositlar xam bo'rva ular bilan quyidagi jadvalda batafsil tanishishingiz mumkin bo'ladi.

Table 4: Machine Learning (ML) Modellarini Yaratish uchun Vositalar

Tur	Vositalar	Misollar
Dasturlash tillari	Python, R, Java, Julia, MATLAB	Python ko'pincha ML uchun ishlatiladi
Kutubxonalar va Frameworklar	Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, Pandas, NumPy	TensorFlow va PyTorch chuqur o'rganish uchun ishlatiladi
Integratsion Rivojlantirish Muhitlari (IDE)	Jupyter Notebook, PyCharm, VS Code, Spyder	Jupyter Notebook ma'lumotlarni tahlil qilish uchun qulaydir
Vizualizatsiya vositalari	Matplotlib, Seaborn, Plotly, Tableau	Matplotlib va Seaborn ma'lumotlarni grafika yordamida tasvirlash uchun ishlatiladi
Bulut platformalari	Google Colab, AWS SageMaker, Microsoft Azure, IBM Watson	Google Colab bepul hisoblash resurslarini taklif etadi

Quyidagi savollarga javob bering va o'tilgan darsdagi bilimlaringizni mustahkamlang

- Classificationga doir o'zingiz 4 ta misol toping va dataset yarating(misoldagidek)
- Regressionga doir o'zingiz 4 ta misol toping va dataset yarating(misoldagidek)
- O'tilgan mavzu bo'yicha tushungalarizni qisqacha bayon qiling
- Takliflar va savollar bering

Classificationga doir o'zingiz misollar va ular uchun dataset.

Mevaning turini aniqlash.

Rangi: qizil, sariq, yashil.
Kattaligi: kichik, o'rta, katta.
Shakli: yumaloq, uzunchoq.

Natija: olma, nok, banan.

Ob-havoni uning darajasiga qarab sinflarga bo'lish:

Harorat qiymati: -15, +6, +25

Natija: sovuq, iliq, issiq

Regressionga doir misollar va ular uchun dataset.

Yillik maoshni prognozlash.

Tajriba: 1yil, 3yil, 5yil, 10yil
Ma'lumoti: bakalavr, magistr, PhD
Soha: IT, tibbiyot, moliya

Natija: 40000\$, 70000\$, 50000\$

Safar vaqtini prognoz qilish.

Masofa: 10km, 50km, 100km.
O'rtacha tezlik: 60km/s, 80km/s, 100km/s

Natija: 15min, 30min, 50min