



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2024"

MASHINALI O'QITISHDA CHIZIQLI REGRESSIYA MASALASI.ENG KICHIK KVADRATLAR USULI

Ilmiy rahbar:

Yaxshiboyev Maxmadiyor Umirovich

Talaba:

Mamirov Ma'rufjon

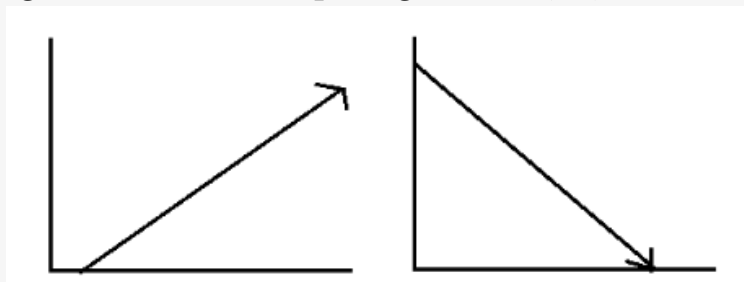
TATU Samarqand filiali

Annotatsiya: Mazkur maqolada, chiziqli regressiyaning nazariy asoslari, eng kichik kvadratlar usulining matematik formulalari va ularning amaliy misollarda qo'llanilishi ko'rib chiqiladi. Shuningdek, modelni baholash va uning samaradorligini oshirish uchun qo'llaniladigan turli texnikalar, jumladan, modelni tasdiqlash (validation), modelni soddalashtirish (regularization), va ma'lumotlar oldindan ishlovi (preprocessing) masalalari yoritiladi.

Kalit so'zlar: Mashinali o'qitish, regressiya, chiziqli regressiya, eng kichik kvadratlar usuli, texnologiyalar, modeling va bashorat qilish, korrelyatsiya koeffitsenti, pandas numpy kutubxonalari

KIRISH

Pythonda chiziqli regressiyaga qisqacha kirish - Chiziqli regressiya ikki (oddiy chiziqli regressiya) yoki undan ortiq (ko'p chiziqli regressiya) o'zgaruvchilar o'rtasidagi chiziqli munosabatni o'rganuvchi statistik model bo'lib, bog'liq bo'lgan o'zgaruvchilardir. o'zgaruvchi va mustaqil o'zgaruvchi (lar).



ko'rib turganingizdek chiziqli regressiya birtamonlama tepaga va bir tamonlama pastga qarab chizilgan. Pythonda regressiya modellarini amalga oshirishga e'tibor qarataman, shuning uchun regressiya chegarasi ostida matematikaga chuqur bog'lanadi.

Pythonda chiziqli regressiya - Pythonda chiziqli regressiyani amalga oshirishning ikkita asosiy usuli bor - Statsmodels va scikit bilan o'rganish. Siz Scikit kutubxonasidan ham foydalanishingiz mumkin. Keling, ikkalasida ham chiziqli regressiyani amalga oshirishni ko'rib chiqaylik:



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2024"

Statsmodelsda chiziqli regressiya - Statsmodels "Ko'p turli statistik modellarni baholash, shuningdek, statistik testlarni o'tkazish va statistikani o'rganish uchun sinflar va funktsiyalarni ta'minlovchi Python moduli." Pandas va NumPy-da bo'lgani kabi, Statsmodels-ni olish yoki o'rnatishning eng oson yo'li Anaconda paketi orqali amalga oshiriladi.

```
import statsmodels.api as sm
```

Chiziqli regressiya uchun Statsmodelsdan qanday foydalanishni ko'rib chiqaylik. Data Science sinfidan misol keltirgan. Birinchidan, biz ma'lumotlar to'plamini sklearn dan import qilamiz.

```
from sklearn import datasets ## imports datasets from scikit-learn
```

```
data = datasets.load_boston() ## loads Boston dataset from datasets library
```

Data.feature_names va data.target ishga tushirilsa, mos ravishda mustaqil o'zgaruvchilar va bog'liq o'zgaruvchilar ustun nomlari chop etiladi. Ya'ni, Scikit-learn allaqachon qiymati va narx ma'lumotlarini maqsadli o'zgaruvchi sifatida va boshqa 13 o'zgaruvchini bashorat qiluvchi sifatida belgilab qo'ygan. Keling, ushbu ma'lumotlar to'plamida chiziqli regressiyani qanday bajarishni ko'rib chiqaylik.

Birinchidan, biz osonroq tahlil qilish uchun ma'lumotlarni pandas dataframe sifatida yuklashimiz va o'rtacha qiymatini maqsadli o'zgaruvchi sifatida belgilashimiz kerak:

```
import numpy as np
```

```
import pandas as pd # define the data/predictors as the pre-set feature names
```

```
df = pd.DataFrame(data.data, columns=data.feature_names)
```

```
# Put the target (housing value -- MEDV) in another DataFrame
```

```
target = pd.DataFrame(data.target, columns=["MEDV"])
```

Bu yerda qilgan ishimiz ma'lumotlar to'plamini olib, uni pandalar ma'lumotlar ramkasi sifatida yuklash; shundan so'ng biz bashorat qiluvchilarni o'rnatamiz (masalan, df) - ma'lumotlar to'plamida oldindan o'rnatilgan mustaqil o'zgaruvchilar. Shuningdek, biz maqsadni qo'ydik - biz taxmin qilmoqchi bo'lgan bog'liq o'zgaruvchi malum bo'ladi.

Keyinchalik, chiziqli regressiya modeliga moslashmoqchi bo'lgan o'zgaruvchilar. o'zgaruvchi uchun yaxshi bashorat qiluvchi o'zgaruvchilarni tanlashi, o'zgaruvchilar orasidagi korrelyatsiyalarni tekshirish, ma'lumotlarning chizmalarini tuzish va aloqalarni vizual ravishda izlash, qaysi o'zgaruvchilar bo'lishi haqida dastlabki tadqiqotlar o'tkazish orqali amalga oshirish mumkin. y uchun yaxshi bashorat qiluvchilar va boshqalar. Ushbu birinchi misol uchun RM ni olaylik - xonalar o'rtacha soni va LSTAT - aholining pastki holatining foizi. Shuni ta'kidlash



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2024"

kerakki, Statsmodels standart konstantani qo'shmaydi. Keling, buni birinchi navbatda regressiya modelimizda doimiysiz ko'rib chiqaylik:

```
## Without a constant
import statsmodels.api as sm
X = df["RM"]
y = target["MEDV"]
# Note the difference in argument order
model = sm.OLS(y, X).fit()
predictions = model.predict(X) # make the predictions by the model
# Print out the statistics
model.summary()
```

SKLearnda chiziqli regressiya - Pythonda mashinani o'rganish haqida gap ketganda, SKLearn deyarli oltin standartdir. U regressiya, tasniflash, klasterlash va o'lchamlarni kamaytirish uchun ko'plab o'rganish algoritmlariga ega. Turli xil algoritmlar xaritasi va SKLearn ga boshqa havolalar uchun KNN algoritmidagi foydalanadi.

Chiziqli regressiyadan foydalanish uchun uni import qilishimiz kerak:

```
from sklearn import linear_model
```

Keling, avvalroq foydalangan ma'lumotlar to'plamidan, uy narxlaridan foydalanaylik. Dastlab, jarayon bir xil bo'ladi: SKLearn-dan ma'lumotlar to'plamlarini import qiling va ularni uy ma'lumotlar to'plamiga yuklang:

```
from sklearn import datasets ## imports datasets from scikit-learn
data = datasets.load_boston() ## loads home dataset from datasets library
Keyinchalik, ma'lumotlarni Pandasga yuklaymiz (avvalgidek):
```

```
# define the data/predictors as the pre-set feature names
df = pd.DataFrame(data.data, columns=data.feature_names)
# Put the target (housing value -- MEDV) in another DataFrame
target = pd.DataFrame(data.target, columns=["MEDV"])
```

Shunday qilib, endi, avvalgidek, bizda mustaqil o'zgaruvchilar ("df") va bog'qa o'zgaruvchiga ega bo'lgan ma'lumotlar to'plami ("maqsad" belgisi) mavjud. SKLearn yordamida regressiya modelini moslashtiramiz. Avval biz X va Y ni aniqlaymiz - bu safar uy narxlarini taxmin qilish uchun dataframedagi barcha o'zgaruvchilardan foydalanamiz:

```
X = df
y = target["MEDV"]
Va yaratilganmodelni olamiz:
lm = linear_model.LinearRegression()
```



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2024"

`model = lm.fit(X,y)`

`lm.fit ()` funksiyasi chiziqli modelga mos keladi. Biz bashorat qilish uchun modeldan foydalanmoqchimiz, Shuning uchun biz `lm.predict ()` dan foydalanamiz:

`predictions = lm.predict(X)`

`print(predictions)[0:5]`

Chop etish funksiyasi `y` uchun dastlabki 5 ta bashoratni chop etadi (men "joyini tejash" uchun butun ro'yxatni chop qilmadim. `[0: 5]` olib tashlansa, butun ro'yxat chop etiladi):

Eng kichik kvadratlar usuli. Regressiya tenglamasi

Ko'pincha ξ tasodifiy miqdor ustida ko'zatish olib borish natijasida hosil qilingan x_1, x_2, \dots, x_n miqdorga boshqa η tasodifiy miqdorning ta'sirini o'rganishga to'g'ri keladi. Agar ξ tasodifiy miqdorning har bir qiymatiga biror qonun asosida η tasodifiy miqdorning aniq qiymati mos kelsa, u holda ξ va η orasidagi munosabat statistik yoki korrelyasion munosabat deyiladi va uni $\eta = \xi(\xi)$ kabi belgilanadi. Bu yerda ξ - munosabat. Aytaylik, ξ va η tasodifiy miqdorlar ustida ko'zatish olib borilgan bo'lib, ko'zatish natijalari mos ravishda x_1, x_2, \dots, x_n , va y_1, y_2, \dots, y_n , lardan iborat bo'lsin, u holda ξ va η orasidagi η - munosabatni 1.1-jadval ko'rinishda ifodalash mumkin:

1.1-jadval.

ξ :	x_1 ,	x_2 ,	...	x_n ,
η :	y_1 ,	y_2 ,	...	y_n ,

Korrelyatsion munosabat to'g'ri, teskari, to'g'ri chiziqli, egri chiziqli va boshqa turlarda bog'langan bo'lishi mumkin. Biz ξ va η orasidagi statistik munosabat chiziqli (to'g'ri chiziqli) bo'lgan holni qaraymiz. Aniqrog'i $\eta = x$, $\xi = y$ deb qarab x ning o'zgarishiga qarab y ni va y ning o'zgarishiga qarab x ni aniqlash masalasini qarab chiqaylik. Bu bog'lanish to'g'ri chiziqli hol bo'lgani uchun $y = kx + b$ ko'rinishda qidirish mumkin. Bu yerda k va b lar noomalum parametrlar bo'lib, ularni "eng kichik kvadratlar usuli" dan foydalanib topamiz. "eng kichik kvadratlar usuli" ga ko'ra, agar y_1, y_2, \dots, y_n ko'zatish natijalaridan iborat bo'lib, bu qiymatlar bilan y ning x_1, x_2, \dots, x_n larga mos keluvchi qiymatlari orasidagi ayirmalar kvadratlarining yig'indisi eng kichik bo'lsa, yaxshi natijaga erishgan bo'ladi.

Shu maqsadda



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2024"

$$F(k, b) = \sum_{i=1}^n (y_{x_i} - y_i)^2 \quad (1)$$

yoki $y_{x_i} = k \cdot x_i + b$ ekanligini e'tiborga olsak,

$$F(k, b) = \sum_{i=1}^n (kx_i + b - y_i)^2 \text{ funksiyani qaraymiz. Bu ifoda eng kichik qiymatga}$$

erishishi uchun

$$\begin{cases} \frac{dF(k, b)}{dk} = 0 \\ \frac{dF(k, b)}{db} = 0 \end{cases}$$

bo'lishi yoki

$$\begin{cases} \sum_{i=1}^n (kx_i + b - y_i)x_i = 0 \\ \sum_{i=1}^n (kx_i + b - y_i) = 0 \end{cases}$$

bajarilishi kerak

Oxirigidan vig'indining xossasiga asosan

$$\begin{cases} k \sum_{i=1}^n x_i^2 + b \cdot \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ k \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \end{cases} \quad (2)$$

kelib chiqadi. Sistemaning har bir tenglamasini n ga bo'lib

$$\begin{cases} k \frac{1}{n} \sum_{i=1}^n x_i^2 + b \cdot \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^n x_i y_i \\ k \frac{1}{n} \sum_{i=1}^n x_i + b = \frac{1}{n} \sum_{i=1}^n y_i \end{cases}$$

ga ega bo'lamiz. Soddaqlik uchun ushbu belgilashlarni kiritamiz:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}, \overline{xy} = \frac{\sum_{i=1}^n y_i x_i}{n}; \bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \overline{X^2} = \frac{\sum_{i=1}^n x_i^2}{n} \quad (3)$$

Bu belgilashlarga nisbatan (3) ni quyidagi ko'rinishda yozish mumkin.

$$\begin{cases} k\bar{X}^2 + b\bar{x} = \overline{xy} \\ k\bar{x} + b = \bar{y} \end{cases} \quad (4)$$

Kelib chiqqan (4) sistemani k va b noma'lumga nisbatan noma'lumni yo'qotish usuli bilan yechamiz. Buning uchun (4) ning ikkinchi tenglamasini \bar{x} ga ko'paytirib, birinchi tenglama bilan hadma-had qo'shib,

$$k\{\bar{x}^2 - [\bar{x}]^2\} = \overline{xy} - \bar{x} \cdot \bar{y} \quad (5)$$

yoki bundan

$$k = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - [\bar{x}]^2} \quad (6)$$

nomalumni topamiz. Topilgan k ning qiymatini (4) ning ikkinchi tenglamasiga qo'ysak

$$\bar{y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - [\bar{x}]^2} \bar{x} + b$$

kelib chiqadi

Bundan

$$b = \bar{y} - \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - [\bar{x}]^2} \bar{x}$$

noma'lumni topamiz. Topilgan k va b parametrlar qiymatini $y_x = kx + b$ ifodaga qo'yib,

Активация Windows
тобы активировать Wind



$$y_x = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - [\bar{x}]^2} \cdot x - \bar{y} - \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - [\bar{x}]^2} \cdot \bar{x}$$

topamiz. Oxirgidan

$$y_x - \bar{y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - [\bar{x}]^2} (x - \bar{x}) \quad (7)$$

kelib chiqadi. (7) ifoda y ning x ga nisbatan regressiya tenglamasi deb ataladi. Xuddi shu kabi yuqoridagi mulohazalarni yuritib va bajarilgan amallarni takrorlab x ning y ga nisbatan regressiya

$$x_y - \bar{x} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - [\bar{x}]^2} (y - \bar{y}) \quad (8)$$

tenglamasini keltirib chiqarish mumkin. Bu yerda $\frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - [\bar{x}]^2}$ nisbatni - bog'liqsiz

miqdor korrelyatsiya koeffitsenti deb ataladi va uni qisqacha r_{xy} kabi belgilanadi.

XULOSA

Xulosa o'rnida aytadigan bo'lsak Chiziqli regressiya ma'lumotlar o'rtasidagi ma'lumotlar yoki aloqani o'rganish uchun foydalaniladigan statistik usuldir. Bu usulda, bir o'zgaruvchili (masalan, X) va uning natijadagi ma'lumotlar (masalan, Y) mavjud bo'lib, ulardagi aloqani ifodalovchi lineyka formulani topishga harakat qilindi. Chiziqli regressiya masalasi asosan quyidagi bosqichlarda hal qilinishini o'rgandik .Ya'ni: Ma'lumotlarni To'plash: Ma'lumotlar to'plami (X va Y) yig'iladi. X o'zgaruvchisi odatda o'zgaruvchilik (muddat, narx, hajm va h.k.) bo'lishi mumkin, Y esa odatda o'zgaruvchiliklar natijasi (masalan, narxni yoki sifatni ifodalaydi).

Modelni Tanlash: Chiziqli regressiya modelini tanlash uchun ma'lumotlardan foydalanish. Model ko'rsatkichlarini (masalan, qarshiliklar va o'lchovchi muhimliklar) aniqlab, barcha ma'lumotlar to'plamini bitta formulaga (masalan, $Y = aX + b$) keltirib chiqamiz.

Modelni Test Qilish: Modelni test qilib, uning aniqlik darajasini va ma'lumotlarni yaxshi aniqlashini baholang.

Natijalarni Tafsilotlash: Model natijalarini tafsilotlash va o'zgaruvchilar bilan aloqani tushunish. Bu qisqacha tushunchalar, chiziqli regressiya usulini ma'lumotlar tahlili, prognostika va boshqa sohalar uchun qo'llashda yordam beradi. Bu usul mashhur machine learning (mashina o'qitish) algoritmlaridan biri bo'lib, sifatlil natijalarni olish va biron-bir tashqi o'zgaruvchilardan kelib chiqqan tushunchalarni sinash uchun qulayliklar keltiradi.



"INNOVATIVE ACHIEVEMENTS IN SCIENCE 2024"

FOYDALANILGAN ADABIYOTLAR:

1. Коломаев А.В. и др. "Теория вероятностей и математическая статистика для экономистов", М. 1991 г.
2. Н.Р.Бекназарова, Х.Н.Жумаев "Математик программалаштириш ва оптималлаштириш" Ўқув предмети бўйича Ўқув-услубий мажмуа (Бакалавриат босқичи талабалари учун). Ташкент 2006.
3. Сафаева К. ва бошқалар. Математик программалашдан маъруза мантлари. Т., ТДМИ, 2003й.
4. В.В.Розен. Математические модели принятия решений в экономике. М. 2002.
5. Математическое программирование в экономике. Под ред. Кремера, М., Финансы и статистика, 1996г.
6. К.Сафаева, Ф.Шомансурова. Математик программалаштиришдан масалалар туплами. Т., Молия институти, 2003й.