

LOGİSTİK REQRESSİYA XƏRÇƏNG ŞİŞİNİN TƏHLİLİ

Bir və ya bir neçə müstəqil dəyişənə malik olan və nəticənin müəyyən edilməsi üçün istifadə olunan statistik idarəetmədir. Mövcud verilənlər toplusunun təhlili iki mümkün nəticədən seçim etməyə imkan verir. Təsnifat məsələlərində istifadə olunur. Logistik reqressiya 1 və ya 0 kimi ikili kodlanmış məlumatları ehtiva edir.

Logistik reqressiya

Logistik reqressiya iki data faktoru arasındakı əlaqələri tapmaq üçün riyaziyyatdan istifadə edən dataların təhlili üsuludur. Logistik reqressiya bu əlaqədən asılı dəyişənə qarşı müstəqil dəyişənə əsaslanan amillərdən birinin dəyərini proqnozlaşdırmaq üçün istifadə edir. Müstəqil dəyişənlərin asılı dəyişənləri var. Bunlar cüt və ya çoxlu qruplaşdırılır. Yeni müstəqil dəyişən gəldikdə onun hansı qrupa aid olduğu müəyyən edilir. Proqnoz adətən məhdud sayda nəticəyə malikdir, bəli və ya xeyr.

Məsələn, deyək ki, siz veb sayt ziyarətçinizin alışı səbətindəki yoxlama düyməsini klikləyəcəyini təxmin etmək istəyirsiniz. Logistik reqressiya təhlili veb saytda sərf olunan vaxt və səbətdəki əşyaların sayı kimi keçmiş ziyarətçi davranışlarına baxır. Tarixən, əgər ziyarətçilər saytda beş dəqiqədən çox vaxt keçiriblərsə və səbətə üçdən çox element əlavə ediblərsə, bu onların yoxlama düyməsini kliklədiklərini müəyyən edir. Bu məlumatdan istifadə edərək, statistik reqressiya funksiyası daha sonra yeni veb-sayt ziyarətçisinin davranışını proqnozlaşdırma bilər.

Logistik reqressiya niyə vacibdir?

Logistik reqressiya süni intellekt və maşın öyrənməsi (AI/ML) sahəsində mühüm texnikadır. ML modelləri (Alqoritmlər) insan müdaxiləsi olmadan mürəkkəb məlumat emal tapşırıqlarını yerinə yetirmək üçün məşq ediləcəyiniz proqram proqramlarıdır. Logistik reqressiyadan istifadə edərək qurulan ML modelləri təşkilatlara biznes məlumatlarından təsirli fikirlər əldə etməyə kömək edir. Onlar əməliyyat xərclərini azaltmaq, səmərəliliyi artırmaq və miqyasını daha sürətli etmək üçün bu məlumatı proqnozlaşdırma təhlil üçün istifadə edilə bilər. Məsələn, müəssisələr işçilərin saxlanması artıran və ya daha sərfəli məhsul dizaynına səbəb olan nümunələri aça bilər.

Logistik reqressiyadan istifadənin digər ML texnikalarından bəzi üstünlükləri bunlardır:

- **Sadəlik:** Logistik reqressiya modelləri digər ML metodlarından daha az riyazi mürəkkəbdir. Buna görə də, komandanızda heç kimin dərin ML təcrübəsi olmasa belə, onları həyata keçirə bilərsiniz.
- **Sürət:** Logistik reqressiya modelləri yaddaş və emal gücü kimi daha az hesablama qabiliyyəti tələb etdiyi üçün böyük həcmli məlumatları yüksək sürətlə emal edə bilir. Bu, onları sürətli qələbələr əldə etmək üçün ML layihələri ilə başlayan təşkilatlar üçün ideal hala gətirir.
- **Görünüş:** Logistik reqressiya təhlili tərtibatçılara digər data təhlili üsullarına nisbətən daxili proqram proseslərinə daha çox görünürlük verir. Problemlərin aradan qaldırılması və səhvlərin düzəldilməsi də daha asandır, çünki hesablamalar daha az mürəkkəbdir.

Logistik reqressiyanın tətbiqləri hansılardır?

Logistik reqressiyanın bir çox müxtəlif sənayelərdə bir neçə real tətbiqi var. İstehsalat: İstehsal şirkətləri maşınlarda hissələrin nasazlığı ehtimalını proqnozlaşdırmaq üçün logistik reqressiya analizindən istifadə edirlər. Sonra gələcək uğursuzluqları minimuma endirmək üçün bu proqnoza əsaslanaraq texniki xidmət cədvəllərini planlaşdırırlar.

Səhiyyə: Tibbi tədqiqatçılar xəstələrdə xəstəlik ehtimalını proqnozlaşdırmaqla profilaktik qayğı və müalicəni planlaşdırırlar. Ailə tarixinin və ya genlərin xəstəliklərə təsirini müqayisə etmək üçün logistik reqressiya modellərindən istifadə edirlər.

Maliyyə: Maliyyə şirkətləri fırldaqçılıq üçün maliyyə əməliyyatlarını təhlil etməli və risk üçün kredit müraciətləri və sığorta müraciətlərini qiymətləndirməlidirlər. Logistik reqressiya modelləri yüksək risk və ya aşağı risk və fırldaqçılıq və ya heç bir saxtakarlıq kimi ayrı nəticələrə malik olduğundan, bu məsələlər logistik reqressiya modeli üçün uyğundur.

Marketing: Onlayn reklam alətləri, istifadəçilərin reklama klik edib-etməyəcəyini proqnozlaşdırmaq üçün logistik reqressiya modelləşdirməsindən istifadə edir. Nəticədə marketoloqlar müxtəlif sözlərə və şəkillərə istifadəçi cavablarını təhlil edə və müştərilərin cəlb edəcəyi yüksək performanslı reklamlar yarada bilərlər.

Reqressiya təhlili necə işləyir?

Logistik reqressiya, data saynstlərin maşın öyrənməsində (ML) tez-tez istifadə etdiyi bir neçə fərqli reqressiya təhlili üsullarından biridir. Logistik reqressiyanı başa düşmək üçün ilk növbədə əsas reqressiya təhlilini başa düşməliyik. Aşağıda reqressiya təhlilinin necə işlədiyini nümayiş etdirmək üçün xətti reqressiya təhlili nümunəsi verilmişdir.

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı

Sualı müəyyənləşdirin: İstənilən məlumat təhlili biznes sualı ilə başlayır. Logistik reqressiya üçün xüsusi nəticələr əldə etmək üçün sualı tərtib etməlisiniz:

- Yağışlı günlər aylıq satışlarımıza təsir edəcəkmi? (bəli və ya yox)
- Müştəri hansı növ kredit kartı fəaliyyəti ilə məşğul olur? (səlahiyyətli, saxta və ya potensial saxtakarlıq)

Tarixi məlumatlar toplanır: Sualı müəyyən etdikdən sonra cəlb olunan data amillərini müəyyən etməlisiniz. Daha sonra bütün faktorlar üçün tarixi məlumatları toplayırsınız. Məsələn, yuxarıda göstərilən ilk suala cavab vermək üçün son üç il ərzində hər ay üçün yağışlı günlərin sayını və aylıq satış datalarını toplaya bilərsiniz.

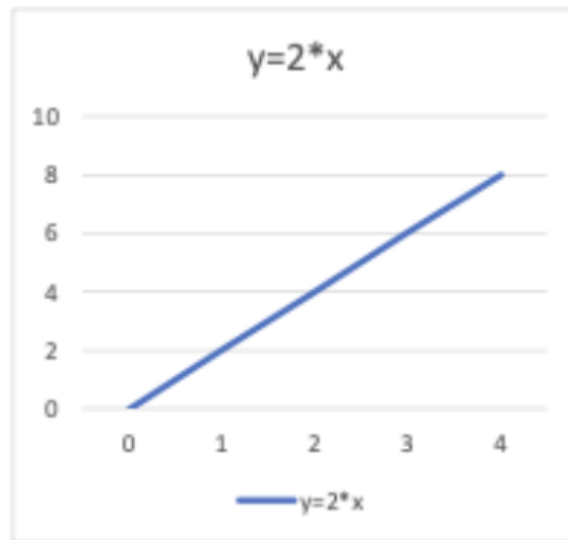
Siz reqressiya təhlili modelini öyrədirsiniz: Siz reqressiya proqramından istifadə edərək tarixi dataları emal edirsiniz. Proqram müxtəlif data nöqtələrini emal edir və tənliklərdən istifadə edərək onları riyazi formada əlaqələndirir. Məsələn, rübdə yağışlı günlərin sayı 3, 5 və 8-dirsə və həmin aylardakı satışların sayı 8, 12 və 18-dirsə, reqressiya alqoritmi amilləri tənlikdə birləşdirəcək: $Satışların sayı = 2 * (Yağışlı Günlərin Sayı) + 2$

Naməlum dəyərlər üçün proqnozlar verir: Naməlum dəyərlər vəziyyətində proqram proqnoz vermək üçün tənlikdən istifadə edir. İyul ayında altı gün yağış olacağını bilirsinizsə, proqram iyulun 14-də satış dəyərini proqnozlaşdıracaq.

Logistik reqressiya modeli necə işləyir?

Logistik reqressiya modelini başa düşmək üçün əvvəlcə tənlikləri və dəyişənləri anlayaq. Tənliklər: Riyaziyyatda tənliklər iki dəyişən arasındakı əlaqəni verir: x və y . Siz x oxu və y oxu boyunca qrafik çəkmək üçün müxtəlif x və y dəyərlərini əvəz etməklə bu tənliklərdən və ya funksiyalardan istifadə edə bilərsiniz. Məsələn, $y = 2 * x$ funksiyasının qrafikini çəksəniz, aşağıda göstəriləndiyi kimi düz xətt alacaqsınız. Buna görə də bu funksiya xətti funksiya da deyilir.

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı



Dəyişənlər: Statistikada dəyişənlər dəyərləri dəyişən data amilləri və ya atributlarıdır. İstənilən təhlil üçün müəyyən dəyişənlər müstəqil və ya izahedici dəyişənlərdir. Bu xüsusiyyətlər nəticənin səbəbidir. Digər dəyişənlər; asılı dəyişənlər və ya cavab dəyişənləridir və onların dəyərləri müstəqil dəyişənlərdən asılıdır. Ümumiyyətlə, logistik reqressiya, hər iki dəyişənin tarixi data dəyərlərinə baxaraq müstəqil dəyişənlərin asılı dəyişənə necə təsir etdiyini araşdırır.

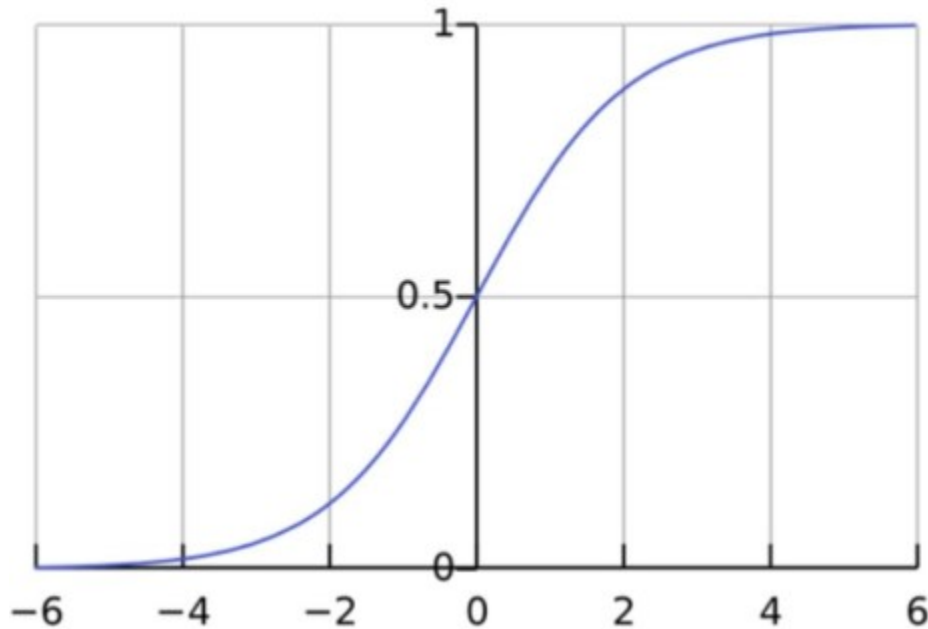
Yuxarıdakı nümunəmizdə x məlum qiymətə malikdir, ona görə də o, müstəqil dəyişən, proqnozlaşdırıcı dəyişən və ya izahedici dəyişən adlanır. Y-nin dəyəri məlum olmadığı üçün ona asılı dəyişən, nəticə dəyişəni və ya cavab dəyişəni deyilir.

Logistik reqressiya funksiyası

Logistik reqressiya riyaziyyatda x və y arasında tənlik kimi logistik funksiya və ya logit funksiyasından istifadə edən statistik modeldir. Logit funksiyası y-ni x-in sigmoid funksiyası kimi təsvir edir.

$$f(x) = \frac{1}{1 + e^{-x}}$$

Bu logistik reqressiya tənliyini tərtib etsəniz, aşağıda göstərildiyi kimi S əyrisi alırsınız.



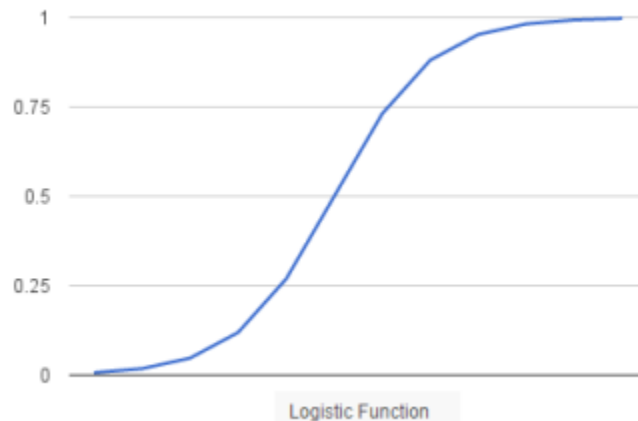
Gördüyünüz kimi, logit funksiyası müstəqil dəyişənin dəyərlərindən asılı olmayaraq asılı dəyişən üçün yalnız 0 ilə 1 arasındakı dəyərləri qaytarır. Logistik reqressiya asılı dəyişənin dəyərini belə proqnozlaşdırır. Logistik reqressiya üsulları həmçinin çoxlu müstəqil dəyişənlər və bir asılı dəyişən arasında tənlikləri modelləşdirir.

Sigmoid funksiyası da adlandırılan logistik funksiya statistiklər tərəfindən sürətlə yüksələrək və tutumunu maksimum dərəcədə artıraraq işlənilib hazırlanmışdır. 0 və 1 arasında bir dəyərə uyğunlaşa bilən hər hansı real dəyərli ədəd S formalı əyridəki dəyərə uyğunlaşdırılır.

Giriş dəyərləri (x) çıxış dəyərini (y) proqnozlaşdırmaq üçün çəkilər və ya əmsal dəyərləri kimi istifadə olunur. Xətti reqressiyadan əhəmiyyətli bir fərq, modelləşdirilmiş çıxış dəyərinin ədədi dəyərdən çox ikili dəyərlər (0 və ya 1) olmasıdır. Aşağıda nümunə logistik reqressiya tənliyi verilmişdir:

$$y = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

burada e natural loqarifmlərin əsasıdır (Euler ədədi və ya EXP() funksiyası) və dəyər çevirmək istədiyiniz faktiki ədədi dəyərdir. Aşağıda logistik funksiya istifadə edərək 0 və 1 diapazonuna çevrilmiş 0 ilə 1 arasındakı ədədlərin qrafiki verilmişdir.



Burada y proqnozlaşdırılan çıxışdır, b_0 proqnoz və ya kəsişmə müddəti, b_1 isə tək giriş dəyəri (x) üçün əmsaldır. Daxil etdiyiniz məlumatların hər bir sütununda train məlumatlarınızdan öyrənilməli olan əlaqəli b əmsalı (sabit faktiki dəyər) var. Yaddaşa və ya faylda saxlayacağınız modelin faktiki təsviri tənlikdəki əmsallardır.

Logistik reqressiya standart sinfin ehtimalını modelləşdirir: Məsələn, əgər biz insanların cinsini boylarına görə kişi və ya qadın kimi modelləşdirsək, logistik reqressiya modeli insanın boyuna görə kişi olma ehtimalı kimi yazıla bilər:

$$P(cins = \text{kişi} | \text{boy})$$

Başqa bir şəkildə yazsaq, girişin (X) standart sinifə ($Y = 1$) aid olması ehtimalını modelləşdiririk, bu belə yazılır:

$$P(X) = P(Y = 1 | X)$$

Ehtimalları proqnozlaşdırmaq? Bununla belə, bu, logistik reqressiyanın təsnifat alqoritmidir. Faktiki olaraq ehtimal qiymətləndirməsi etmək üçün ehtimal təxmini ikili dəyərlərə (0 və ya 1) çevrilməlidir. Logistik reqressiya xətti metoddur, lakin qiymətlər logistik funksiyaadan istifadə edərək çevrilir. Bunun təsiri ondan ibarətdir ki, biz proqnozları xətti reqressiya ilə, məsələn, yuxarıdan davam etdirdiyimiz kimi girişlərin xətti birləşməsi kimi başa düşə bilmirik, model belə ifadə edilə bilər:

$$p(X) = e^{(b_0 + b_1 * X)} / (1 + e^{(b_0 + b_1 * X)})$$

Riyaziyyata çox dərindən girmək istəmirəm, lakin yuxarıdakı tənliyi aşağıdakı kimi fırlada bilərik (qeyd edək ki, e -ni bir tərəfdən təbii loqarifmanı (\ln) digərinə əlavə etməklə çıxara bilərik):

$$\ln(p(X) / (1 - p(X))) = b_0 + b_1 * X$$

Sol tərəf nisbəti standart sinfin ehtimalları adlanır. Nisbətlər hadisənin baş vermə ehtimalını heç bir hadisənin olmama ehtimalına bölmək yolu ilə hesablanır, məs. $0,8 / (1 - 0,8)$ olan 4 ehtimalı var. Beləliklə, bunun əvəzinə yaza bilərik:

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı

$$\ln(\text{nisbətlər}) = b_0 + b_1 * X$$

Nisbətlər log-a çevrildiyindən, sol tərəfə log-egtimal deyilir. Eksponenti sağa qaytarıb yazı bilərik:

$$\text{nisbət} = e^{(b_0 + b_1 * X)}$$

Bütün bunlar, əslində, modelin hələ də girişlərin xətti kombinasiyası olduğunu başa düşməyə kömək edir, lakin bu xətti kombinasiya standart sinfin log-əməlləri ilə bağlıdır.

Logistik regressiya alqoritminin əmsalları (Beta dəyərləri b) train datalarınızdan qiymətləndirilməlidir. Bu, maksimum ehtimal qiymətləndirməsindən istifadə etməklə edilir. Logistik regressiya alqoritminin əmsalları (Beta dəyərləri b) train datalarınızdan qiymətləndirilməlidir. Bu, maksimum ehtimal qiymətləndirməsindən istifadə etməklə edilir. Maksimum ehtimalın qiymətləndirilməsi müxtəlif maşın öyrənmə alqoritmləri tərəfindən istifadə edilən ümumi öyrənmə alqoritmidir, lakin datalarınızın paylanması ilə bağlı fərziyyələr irəli sürür (bu barədə daha çox datalarınızı hazırlamaq haqqında danışarkən).

Ən yaxşı əmsallar, fərz edilən sinif (məsələn, kişi) üçün 1-ə çox yaxın bir dəyər və digər sinif (məsələn, qadın) üçün 0-a çox yaxın bir dəyər proqnozlaşdıran bir modellə nəticələnir. Logistik regressiya üçün maksimum ehtimal intuisiyası ondan ibarətdir ki, axtarış proseduru verilənlərdə olanlarla model tərəfindən proqnozlaşdırılan ehtimallardakı xətanı minimuma endirən əmsallar (beta dəyərləri) üçün dəyərlər axtarır (məsələn, əgər data əsasdırsa, 1 ehtimalıdır).

Biz maksimum ehtimal olunan riyaziyyata girməyəcəyik. Train datalarınız üçün əmsallar üçün ən yaxşı dəyərləri optimallaşdırmaq üçün minimuma endirmə alqoritmindən istifadə edildiyini söyləmək kifayətdir. Bu, adətən, səmərəli ədədi optimallaşdırma alqoritmindən istifadə etməklə praktikada həyata keçirilir (məsələn, Kvazi-Nyuton metodu).

Logistikani öyrənərkən, daha sadə gradient eniş alqoritmindən istifadə edərək bunu özünüz həyata keçirə bilərsiniz.

Çox müstəqil dəyişən logistik regressiya təhlili

Əksər hallarda birdən çox izahedici dəyişən asılı dəyişənin dəyərinə təsir göstərir. Logistik regressiya düsturları bu cür giriş data dəstlərini modelləşdirmək üçün müxtəlif müstəqil dəyişənlər arasında xətti əlaqəni nəzərdə tutur. Sigmoid funksiyasını dəyişdirir və son çıxış dəyişənini bu şəkildə işləyə bilərsiniz:

$$y = f(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)$$

β simvolu reqressiya əmsalını təmsil edir. Logit modeli, həm asılı, həm də müstəqil dəyişənlərin məlum dəyərləri ilə kifayət qədər böyük bir eksperimental data dəstini verdiyiniz zaman bu əmsal dəyərlərini tərsinə hesablaya bilər.

Loqarifmik ehtimallar

Logit modeli həm də müvəffəqiyyətin uğursuzluğa nisbətini və ya loqarifmik əmsalları müəyyən edə bilər. Məsələn, dostlarınızla poker oynayarkən 10 matçdan dördünü qazansanız, udmaq şansınız altıdan dördüdür, bu da uğurunuzun uğursuzluğa nisbətidir. Digər tərəfdən, qazanma ehtimalınız 10-da dördüdür.

Riyazi olaraq ehtimal baxımından əmsallarınız $p/(1 - p)$ və loqarifmik əmsallarınız $(p/(1 - p))$ təşkil edir. Logistik funksiyanı aşağıda göstərilədiyi kimi loqarifmik ehtimallar kimi təqdim edə bilərsiniz:

$$\text{Logit Funksiyası} = \log \left(\frac{p}{1 - p} \right)$$

Asılı dəyişənin nəticələrinə əsasən logistik reqressiya təhlilinə üç yanaşma mövcuddur.

- Binar logistik reqressiya: Binar logistik reqressiya yalnız iki mümkün nəticəsi olan ikili təsnifat problemləri üçün işləyir. Asılı dəyişənin yalnız iki dəyəri ola bilər, məsələn, "bəli və yox" və ya "0 və 1". Logistik funksiya 0 ilə 1 arasında bir sıra dəyərlər hesablasa da, ikili reqressiya modeli cavabı ən yaxın dəyərə yuvarlaqlaşdırır. Əsasən 0.5-dən aşağı olan cavablar 0-a yuvarlaqlaşdırılır və 0.5-dən yuxarı cavablar 1-ə yuvarlaqlaşdırılır; belə ki, logistik funksiya ikili nəticə qaytarır.
- Multinomial logistik reqressiya: Çoxnomlu reqressiya, nəticələrin sayı məhdud olduğu müddətdə bir neçə mümkün nəticə ilə problemləri təhlil edə bilər. Məsələn, əhalinin məlumatlarına əsasən ev qiymətlərinin 25%, 50%, 75% və ya 100% artacağını proqnozlaşdırmaqla bəli, lakin evin dəqiq dəyərini proqnozlaşdırmaqla bilməz. Multinomial logistik reqressiya nəticə dəyərlərini 0 ilə 1 arasında diskret dəyərlərə uyğunlaşdırmaqla işləyir. Logistik funksiya 0.1; 0.11; 0.12 və s. kimi bir sıra davamlı verilənləri qaytara bildiyi üçün, multinomial reqressiya da çıxışı mümkün olan ən yaxın dəyərlərlə qruplaşdırır.
- Ardıcıl logistik reqressiya: ardıcıl logistik reqressiya və ya ardıcıl logit modeli, ədədlərin həqiqi dəyərləri deyil, sıralamanı təmsil etdiyi tapşırıqlar üçün xüsusi bir polinom reqressiya növüdür. Məsələn, müştərilərin il ərzində sizdən aldığı məhsulların sayı kimi ədədi dəyərə əsaslanaraq xidmətinizi pis, orta, yaxşı və ya əla

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı

kimi qiymətləndirmələrini xahiş edən sorğu sualına cavablarını proqnozlaşdırmaq üçün sıralanmış reqressiyadan istifadə edirsiniz.

Logistik reqressiya digər MÖ texnikaları ilə müqayisədə necə işləyir?

İki ümumi data təhlili texnikası; xətti reqressiya təhlili və dərin öyrənmə.

Xətti reqressiya təhlili

Yuxarıda izah edildiyi kimi, xətti reqressiya xətti birləşmədən istifadə edərək asılı və müstəqil dəyişənlər arasındakı əlaqəni modelləşdirir. Xətti reqressiya tənliyi

$$y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n + \epsilon, \text{ burada } \beta_1 \text{ ilə } \beta_n \text{ və } \epsilon \text{ reqressiya əmsalıdır.}$$

Logistik reqressiyanın və xətti reqressiyanın müqayisəsi

Xətti reqressiya verilmiş müstəqil dəyişənlər toplusundan istifadə edərək davamlı asılı dəyişəni proqnozlaşdırır. Davamlı dəyişən qiymət və ya yaş kimi bir sıra dəyərlərə malik ola bilər. Beləliklə, xətti reqressiya asılı dəyişənin həqiqi dəyərlərini proqnozlaşdırmağa bilər. "10 ildən sonra düynünün qiyməti nə qədər olacaq?" kimi suallara cavab verə bilər.

Xətti reqressiyadan fərqli olaraq, logistik reqressiya təsnifat alqoritmidir. Kəsilməz datalar üçün həqiqi dəyərləri proqnozlaşdırmağa bilməz. "10 ildən sonra düynünün qiyməti 50% artacaqmı?" kimi suallara cavab verə bilər.

Dərin Öyrənmə

Dərin öyrənmə məlumatı təhlil etmək üçün insan beynini simulyasiya edən neyron şəbəkələrdən və ya proqram komponentlərindən istifadə edir. Dərin öyrənmə hesablamaları vektorların riyazi konsepsiyasına əsaslanır.

Logistik reqressiya və dərin öyrənmə müqayisəsi

Logistik reqressiya dərin öyrənmə ilə müqayisədə daha az mürəkkəb və daha az hesablama intensivliyinə malikdir. Daha da əhəmiyyətli, dərin öyrənmə hesablamaları mürəkkəbliyi və maşın yönümlü təbiətinə görə tərtibatçılar tərəfindən araşdırıla və ya dəyişdirilə bilməz. Digər tərəfdən, logistik reqressiya hesablamaları şəffafdır və problemləri həll etmək daha asandır.

Logistik reqressiya modelini öyrənmək:

Logistik reqressiya modeli ilə proqnozlaşdırmaq logistik reqressiya tənliyinə ədədlər əlavə etmək və nəticəni hesablamaq qədər sadədir.

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı

Bunu konkret misalla konkretləşdirək. Deyək ki, insanın boyuna görə kişi və ya qadın olduğunu təxmin edə bilən (sırf xəyali) bir modelimiz var. Boy 150 sm verildikdə kişi və ya qadındır. Biz $b_0 = -100$ və $b_1 = 0,6$ əmsallarını öyrəndik.

Yuxarıdakı tənlikdən istifadə edərək, rəsmi olaraq $P(\text{kişi} \mid \text{boy} = 150)$ olan 150 sm və ya daha çox hündürlüyü verilmiş bir kişinin ehtimalını hesablaya bilərik.

$$y = e^{(b_0 + b_1 * X)} / (1 + e^{(b_0 + b_1 * X)})$$
$$y = \exp(-100 + 0.6 * 150) / (1 + \exp(-100 + 0.6 * X))$$
$$y = 0,0000453978687$$

Adamın kişi olma ehtimalı sıfıra yaxındır. Praktikada ehtimallardan birbaşa istifadə edə bilərik. Bu təsnifat olduğundan və aydın cavab istədiyimiz üçün ehtimalları binar sinif dəyərinə əsaslandırma bilərik, məsələn: 0 əgər $p(\text{kişi}) < 0.5$, 1 əgər $p(\text{kişi}) \geq 0.5$

İndi logistik regressiyadan istifadə edərək proqnozlar verməyi bildiyimizə görə, texnikadan maksimum yararlanmaq üçün məlumatlarımızı necə hazırlaya biləcəyimizə baxaq.

Logistik regressiya üçün məlumatların hazırlanması:

Verilənlərinizdə paylanma və əlaqələr haqqında logistik regressiyanın irəli sürdüyü fərziyyələr əsasən xətti regressiyada edilən fərziyyələrlə eynidir. Ciddi ehtimal və statistik dildən istifadə edərək, bu fərziyyələri müəyyən etmək üçün çox iş aparılmışdır. Nəhayət, proqnozlaşdırıcı modelləşdirmə maşın öyrənmə layihələrində siz nəticələri şərh etməkdənsə, dəqiq proqnozlar verməyə diqqət yetirirsiniz.

Buna görə də, model möhkəm olduğu və yaxşı performans göstərdiyi müddətcə bəzi fərziyyələri poza bilərsiniz.

Binar Çıxış Dəyişən: Logistik regressiya ikili (iki sinifli) təsnifat problemləri üçün nəzərdə tutulmuşdur. O, nümunənin standart sinifə aid olma ehtimalını təxmin edəcək və 0 və ya 1 təsnifatına daxil edilə bilər.

Səs-küyü aradan qaldırın: Logistik regressiya çıxış dəyişənində (y) heç bir xətanın olmadığını güman edir, train məlumatlarınızdan kənar göstəriciləri və bəlkə də yanlış təsnif edilmiş nümunələri silməyi düşünün.

Gauss Distribution: Logistik regressiya xətti alqoritmidir (çıxışda qeyri-xətti çevrilmə ilə). O, giriş dəyişənləri və çıxış arasında xətti əlaqəni nəzərdə tutur. Bu xətti əlaqəni daha yaxşı üzə çıxaran giriş dəyişənlərinizin məlumat çevrilmələri daha dəqiq modellə nəticələnmə bilər.

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı

Məsələn, bu əlaqəni daha yaxşı aşkar etmək üçün log, root, BoxCox və digər birdəyişənli çevrilmələrdən istifadə edə bilərsiniz.

Əlaqəli Girişləri Silin: Xətti reqressiya kimi, çoxlu yüksək korrelyasiyalı girişləriniz varsa, model həddindən artıq uyğunlaşa bilər. Bütün girişlər arasında cüt korrelyasiyaları hesablamağı və yüksək korrelyasiyalı qeydləri silməyi nəzərdən keçirin.

Failure to Converge: Əmsalları öyrənən gözlənilən ehtimal qiymətləndirmə prosesinin yaxınlaşmaması mümkündür. Bu, məlumatlarınız çoxlu yüksək korrelyasiyalı girişlərə malik olduqda və ya məlumatlar çox seyrək olduqda baş verə bilər (məsələn, giriş datalarınızda çoxlu sıfırlar).

Nümunə olaraq : Logistik Reqressiya

Logistik reqressiya təsnifat problemlərini həll etmək məqsədi daşıyır. O, bunu davamlı nəticəni proqnozlaşdıran xətti reqressiyadan fərqli olaraq, kateqoriyalı nəticələri proqnozlaşdırmaqla edir. Ən sadə halda, binomials adlanan iki nəticə var; buna misal olaraq şişin bədxassəli və ya xoşxassəli olduğunu proqnozlaşdırmaq olar. Digər şərtlər təsnif edilməli olan ikidən çox nəticəyə malikdir, bu vəziyyət multinomial adlanır. Multinomial logistik reqressiyanın ümumi nümunəsi 3 müxtəlif növ arasında iris çiçəyinin növünün proqnozlaşdırılmasıdır. Burada binomial dəyişəni qiymətləndirmək üçün əsas logistik reqressiyadan istifadə edəcəyik. Bu o deməkdir ki, yalnız iki mümkün nəticə var.

Bu necə işləyir?

Python-da bizim üçün işi görəcək modullarımız var. NumPy modulunu import etməklə başlayaq.

Müstəqil dəyişənləri X-də qeyd edək.

Asılı dəyişəni y-də qeyd edək.

Aşağıda nümunə verilənlər bazası verilmişdir:

#X santimetrlə bir şişin ölçüsünü təmsil edir.

#Qeyd: LogisticRegression() funksiyasının işləməsi üçün X sətirdən sütuna çevrilməlidir.

```
import numpy
```

```
# Qeyd: LogisticRegression() funksiyasının işləməsi üçün X sətirdən sütuna çevrilməlidir.
```

```
X = numpy.array([3.78, 2.44, 2.09, 0.14, 1.72, 1.65, 4.92, 4.37, 4.96, 4.52, 3.69, 5.88]).reshape(-1,1)
```

```
#y şişin xərçəngli olub-olmadığını ifadə edir ("Xeyr" üçün 0, "Bəli" üçün 1).
```

```
y = numpy.array([0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1])
```

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı

Biz sklearn modulundan bir metoddan istifadə edəcəyik, ona görə də həmin modulu da import etməli olacağıq:

```
from sklearn import linear_model
```

Sklearn modulundan logistik reqressiya obyektini yaratmaq üçün LogisticRegression() metodundan istifadə edəcəyik.

Bu obyektə müstəqil və asılı dəyərləri parametr kimi qəbul edən və reqressiya obyektini əlaqəni təsvir edən verilənlərlə dolduran fit() adlı metod var:

```
logr = linear_model.LogisticRegression()  
logr.fit(X,y)
```

```
▼ LogisticRegression  
LogisticRegression()
```

İndi bizdə şiş ölçüsünə əsasən şişin xərçəngli olub-olmamasına hazır olan logistik reqressiya obyektimiz var:

Ölçüsü 3,46 mm olan şişin xərçəngli olub olmadığını təxmin edək:

```
predicted = logr.predict(numpy.array([3.46]).reshape(-1,1))
```

Aşağıda nümunə data seti verilmişdir:

```
#X santimetrlə bir şişin ölçüsünü təmsil edir.
```

```
X = numpy.array([3.78, 2.44, 2.09, 0.14, 1.72, 1.65, 4.92, 4.37, 4.96, 4.52, 3.69, 5.88]).reshape(-1,1)
```

```
#Qeyd: LogisticRegression() funksiyasının işləməsi üçün X cərgədən sütuna çevrilməlidir.
```

```
y = numpy.array([0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1])
```

```
# Biz Sklearn modulundan bir metoddan istifadə edəcəyik, ona görə də həmin modulu da import etməliyik
```

```
from sklearn import linear_model
```

```
# Sklearn modulundan logistik reqressiya obyektini yaratmaq üçün LogisticRegression() metodundan istifadə edəcəyik.
```

```
# Bu obyektə müstəqil və asılı dəyərləri parametr kimi qəbul edən və reqressiya obyektini əlaqəni təsvir edən verilənlərlə
```

```
# dolduran fit() adlı metod var:
```

```
logr = linear_model.LogisticRegression()
```

```
logr.fit(X,y)
```

```
# İndi bir şişin xərçəngli olub-olmadığını, şiş ölçüsünə əsaslanaraq müəyyən etməyə hazır olan logistik reqressiya obyektimiz  
# var.
```

```
# Ölçüsü 3,46 mm olan şişin xərçəngli olub-olmadığını təxmin edək:
```

```
predicted = logr.predict(numpy.array([3.46]).reshape(-1,1))
```

```
print(predicted)
```

```
[0]
```

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı

Ehtimal

Əmsal və kəsişmə dəyərlərindən hər bir şüşin xərcəng olma ehtimalını tapmaq üçün istifadə edilə bilər. Yeni bir dəyər qaytarmaq üçün modelin əmsalı və kəsişmə dəyərlərindən istifadə edən bir funksiya yaradılır. Bu yeni dəyər verilən müşahidənin şüş olması ehtimalını əks etdirir.

```
def logit2prob(logr,x):  
    log_odds = logr.coef_ * x + logr.intercept_  
    odds = numpy.exp(log_odds)  
    probability = odds / (1 + odds)  
    return(probability)
```

Funksiya təsviri: Hər bir müşahidə üçün log-əməlləri tapmaq üçün əvvəlcə əmsalı və kəsişməni çıxararaq xətti reqressiyadan əldə edilənə bənzər bir düstur yaratmaq lazımdır.

```
log_odds = logr.coef_ * x + logr.intercept_
```

Daha sonra log əmsalları əmsallara çevrilmək üçün eksponentləşdirilir.

```
odds = numpy.exp(log_odds)
```

İndi əmsallarımız olduğuna görə, 1 plus əmsallara bölmək yolu ilə onu ehtimala çevirə bilərik.

```
probability = odds / (1 + odds)
```

İndi hər bir şüşin xərcəngli olması ehtimalını tapmaq üçün öyrəndiklərimizlə funksiya istifadə edək.

Gəlin bütün nümunəni icra edək.

MÜƏLLİF: İBRAHİM ABBASZADƏ
MÜƏLLİF HUQUQLARI QORUNUR !
28 Noyabr 2024, cümə axşamı

```
import numpy
from sklearn import linear_model
X = numpy.array([3.78, 2.44, 2.09, 0.14, 1.72, 1.65, 4.92, 4.37, 4.96, 4.52, 3.69, 5.88]).reshape(-1,1)
y = numpy.array([0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1])
logr = linear_model.LogisticRegression()
logr.fit(X,y)

def logit2prob(logr, X):
    log_odds = logr.coef_ * X + logr.intercept_
    odds = numpy.exp(log_odds)
    probability = odds / (1 + odds)
    return(probability)
print(logit2prob(logr, X))
```

[[0.60749955]
[0.19268876]
[0.12775886]
[0.00955221]
[0.08038616]
[0.07345637]
[0.88362743]
[0.77901378]
[0.88924409]
[0.81293497]
[0.57719129]
[0.96664243]]

Nəticələr

3.78, 0.61 — — 3.78 sm ölçülü bir şişin xərçəng olma ehtimalı 61% təşkil edir.
2.44, 0.19 — — 2.44 sm ölçülü şişin xərçəng olma ehtimalı 60% təşkil edir.
2.09, 0.13 — — 2.09 sm ölçülü olan şişin xərçəng olma ehtimalı 13% təşkil edir.