Kredit riskinin modelləşdirilməsi: PD modelləşdirməsində daralma üsulları və LASSO seçimi (2-ci hissə)

written by Hikmat Abdulazizov Hikmət əbdüləzizov

Statistik modelləşdirmə və maşın öyrənməsi sahəsində reqressiya təhlili dəyişənlər arasındakı münasibətləri anlamaq və proqnozlaşdırmaq üçün əsas vasitədir. Ənənəvi xətti reqressiya güclü olsa da, yüksək ölçülü məlumat dəstləri ilə işləyərkən və ya proqnozlaşdırıcılar arasında multikollinearlıq mövcud olduqda tez-tez problemlərlə üzləşir. Belə hallarda, Lasso və Ric reqressiyası kimi üsullar nizamlanmanı tətbiq etməklə əla həllər təklif edir.

Lasso (ən Az Mütləq Büzülmə və Seçmə Operatoru) və Ric reqressiyası həm həddən artıq uyğunlaşmanın azaldılmasına, həm də reqressiya modellərinin ümumiləşdirilməsinin təkmilləşdirilməsinə yönəlmiş üsullardır. Onlar buna reqressiya dəyişənlərinin əmsallarına cərimələr tətbiq etməklə nail olurlar və bununla da onların proqnozlaşdırıcı gücünü qoruyub saxlayaraq onları sıfıra endirirlər. Bu üsullar iqtisadiyyat, genetika və maşın öyrənməsi kimi müxtəlif sahələrdə geniş şəkildə tətbiq edilmişdir.

Lasso və Ric reqressiyası maşın öyrənməsi alət dəstində əvəzolunmaz alətlərdir, çünki onlar modelin mürəkkəbliyinə nəzarət etmək, ümumiləşdirmə performansını artırmaq və xüsusiyyət seçimini asanlaşdırmaq üçün güclü mexanizmlər təklif edirlər. Tənzimləmə prinsiplərindən istifadə etməklə, bu üsullar maşın öyrənməsi ilə məşğul olan mütəxəssislərə geniş tətbiqlərdə daha möhkəm, şərh edilə bilən və proqnozlaşdırıla bilən modellər yaratmağa imkan verir.

Bu yazıda biz Lasso və Ric reqressiya üsullarının prinsiplərini, metodologiyalarını və tətbiqlərini araşdırırıq. Biz onların nəzəri əsaslarını araşdırırıq, fərqlərini və oxşarlıqlarını müzakirə edirik və onların həyata keçirilməsi üçün praktiki mülahizələri təhlil edirik. Bundan əlavə, biz Lasso və Ric reqressiyasının üstün olduğu ssenariləri vurğulayırıq və ənənəvi xətti reqressiya metodları ilə müqayisədə onların üstünlükləri və məhdudiyyətləri haqqında məlumat veririk.

Lasso və Ric reqressiyasının hərtərəfli tədqiqi vasitəsilə biz oxucuları tənzimləmə üsulları və onların müasir məlumatların təhlilindəki rolu haqqında daha dərindən başa salmağı məqsəd götürmüşük. Bu metodların incəliklərini aydınlaşdıraraq, biz tədqiqatçılara, analitiklərə və praktikantlara mürəkkəb real dünya məlumat dəstləri qarşısında dayanıqlı və şərh edilə bilən reqressiya modelləri yaratmaq üçün dəyərli alətlərlə işləmək gücü verməyə çalışırıq.

Lasso və Ric reqressiyası, hər biri fərqli xüsusiyyətlərə və üstünlüklərə malik olan tənzimləmə üçün iki görkəmli yanaşmadır. 1996-cı ildə Robert Tibşirani tərəfindən təqdim edilən Lasso reqressiyası dəyişən seçimini effektiv həyata keçirərək bəzi reqressiya əmsallarını tam olaraq sıfıra endirərək seyrəkliyi təşviq edən L1 cəza həddini özündə birləşdirir. Digər tərəfdən, 1970-ci ildə Hoerl və Kennard tərəfindən təklif edilən Ric reqressiyası, əmsalları tamamilə aradan qaldırmadan sıfıra endirmək üçün L2 cəza həddindən istifadə edir, beləliklə, multikollinearlığın təsirini azaldır və modeli sabitləşdirir.

Müxtəlif fənlər üzrə Lasso və Ric reqressiya üsullarının geniş tətbiqi onların müasir məlumatların təhlilində aktuallığını və effektivliyini vurğulayır. Bu üsullar yalnız ümumi reqressiya problemlərinə etibarlı həllər təklif etmir, həm də xüsusiyyətin əhəmiyyəti və modelin şərh oluna bilməsi haqqında dəyərli fikirlər təqdim edir. Bundan əlavə, hesablama alqoritmlərində və optimallaşdırma üsullarında irəliləyişlər

Lasso və Ric reqressiyasının həyata keçirilməsini əvvəlkindən daha əlçatan və səmərəli etmişdir.

Biz Lasso və Ric reqressiyasının hesablama aspektlərini, o cümlədən parametrlərin qiymətləndirilməsi və model seçimi üçün alqoritmləri müzakirə edirik. Biz həmçinin xüsusiyyət seçimi, proqnozlaşdırıcı modelləşdirmə və risk təhlili kimi sahələrdə bu texnikaların effektivliyini nümayiş etdirən real həyat nümunələri və empirik tapıntıları vurğulayırıq.

Ric reqressiyası

ən sadə halda, singular moment matrisinin problemi,

$$\mathbf{X}^\mathsf{T}\mathbf{X}$$

diaqonallara müsbət elementlər əlavə etməklə yüngülləşir, bununla da onun şərt sayı azalır, daha asan həll oluna bilənə çevrilir. Adi ən kiçik kvadratlar qiymətləndiricisinin analoqu kimi, sadə Ric qiymətləndiricisi aşağıdakı kimi verilir:

$$\hat{\beta}_R = \left(\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}$$

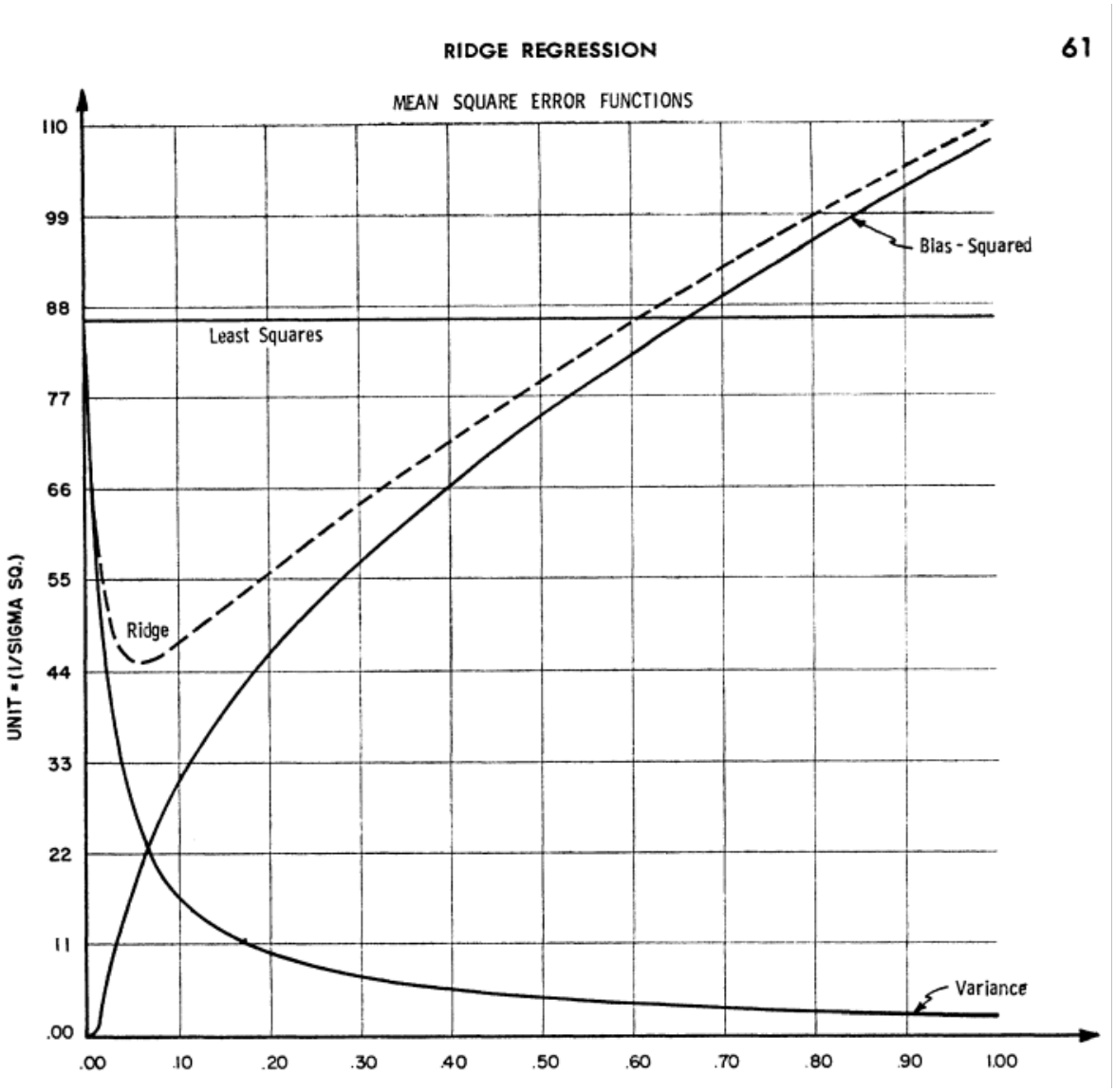
burada y reqress olunan və X dizayn matrisidir, I eynilik matrisidir və Ric parametri müsbət lambda moment matrisinin diaqonallarını dəyişdirən sabit kimi xidmət edir. Qeyd edək ki, lambda sıfır olduqda, tənlik sadə OLS reqressiyasına qədər enir. Aşağıda nizamlanmanın minimuma endirilməsi probleminə baxın.

$$\min_{\beta} \left(\mathbf{y} - \mathbf{X}\beta\right)^\mathsf{T}\left(\mathbf{y} - \mathbf{X}\beta\right) + \lambda\left(\beta^\mathsf{T}\beta - c\right)$$

Birinci hissə sadəcə standart şərti OLS hissəsidir, ikinci hissə isə silsilənin cəzalandırılması üçündür. Buna görə də OLS-dən daha kiçik əmsallarımız var. “Büzülmə modeli” termini

buradan yaranır. OLS əvəzinə silsilənin reqressiyasından istifadə etməyin üstünlüyü aşağıdakı qrafikdə göstərilmişdir.

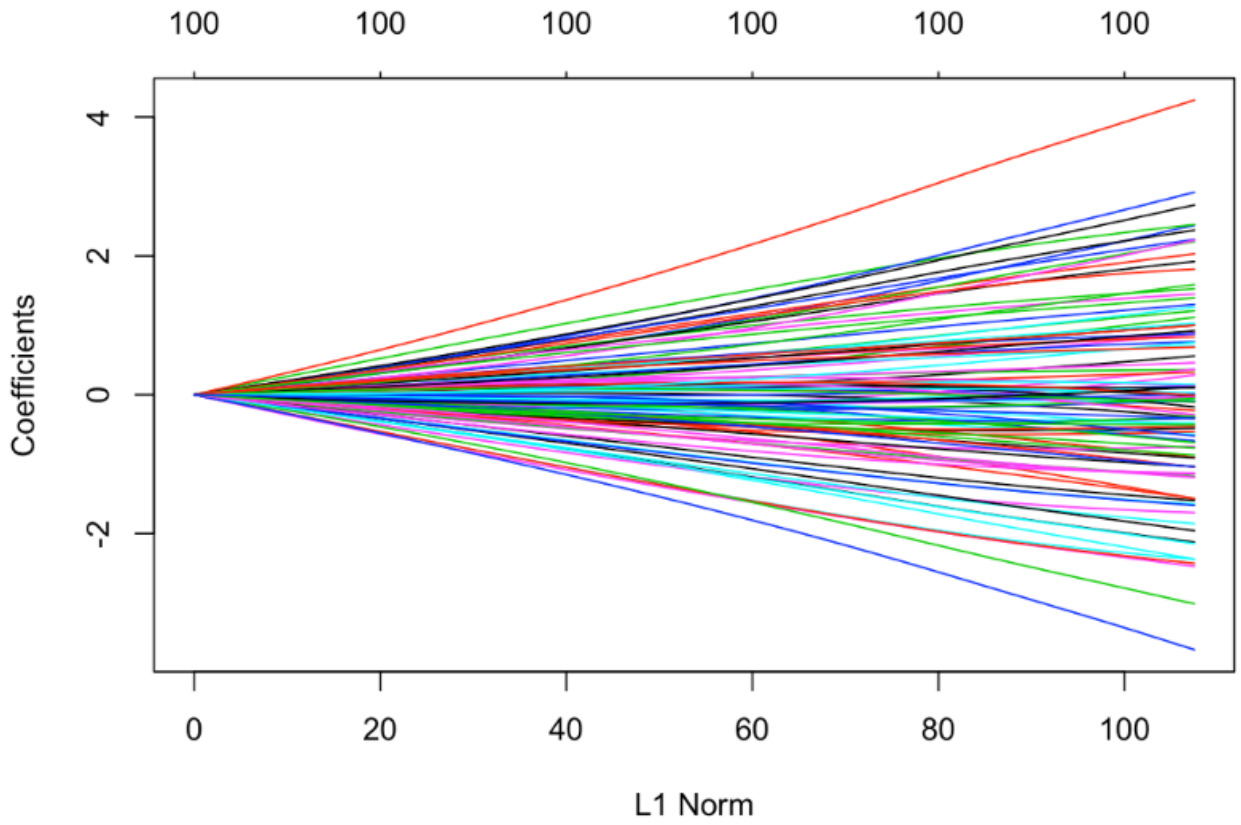
Qrafik 1. Büzülmənin Optimallığı



Yuxarıdakı qrafikdən göründüyü kimi, daha çox meylilik tətbiq etməklə biz dispersiyanı əhəmiyyətli dərəcədə azalda və ümumi orta kvadrat xətanı minimallaşdıra bilərik, çünki orta kvadrat xəta dispersiya və qərəmeyliliyin kvadratının cəmidir. Bu çox vacibdir, çünki bu Ric qiymətləndiricisindən istifadə etməklə biz proqnozlaşdırma gücümüzü artıra və sonda daha etibarlı modellərə sahib ola bilməyimiz deməkdir. İndi simulyasiya ilə glmnet paketindən və real dünya kredit riski məlumatlarından istifadə edərək, Ric modelinin R proqramından çıxan

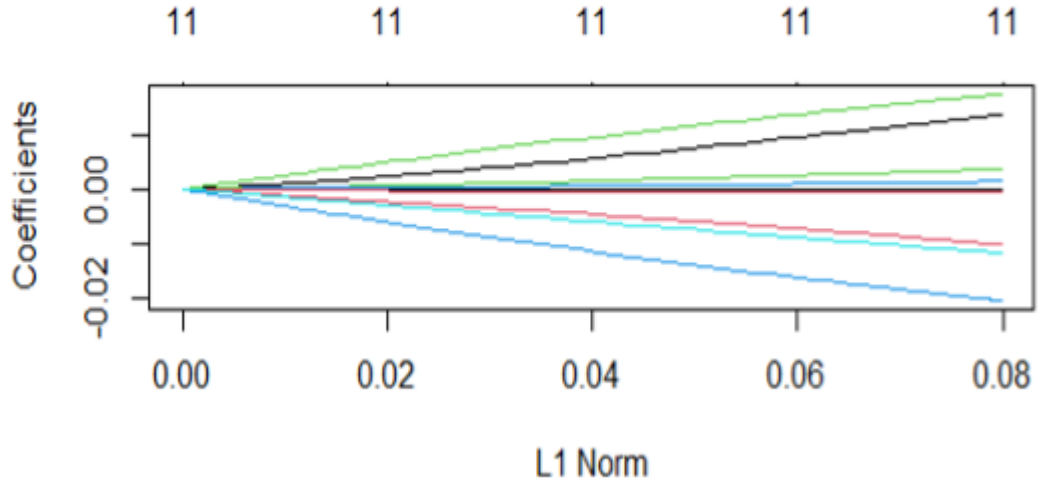
nəticələrinə baxaq.

Qrafik 2. Simulyasiya Edilmiş Ricin əmsal Yolu



Yuxarıda lambda parametrini tənzimləməklə əmsal yollarını göstərən Ric reqressiyasının çıxışı var. Unutmayın ki, lambda artdıqca cəzalandırma şiddətlənir və beta parametrləri daha da kiçilir. Beləliklə, qrafik tənzimləmə parametri azalarkən əmsal yollarını göstərir. Lambda ən yüksək olduqda əmsallar sıfırdan başlayır. Sonra model birdən bütün dəyişənləri götürür (burada 100 dəyişən) və əmsallar artır. Bu, Ric reqressiyasının çox vacib xüsusiyyətidir, çünki o, modeldəki bütün dəyişənləri onların qarşılıqlı əlaqəsindən asılı olmayaraq istifadə edə biləcəyimizi iddia edir. Ən radikal halda mən eyni dəyişəndən istifadə etmişəm (beləliklə, korrelyasiya əmsalı tam olaraq 1-ə bərabərdir) və model hələ də onların hər ikisini seçir.

Qrafik 3. Real Dünya Kredit Datasının Ric əmsalları



Yuxarıdakı qrafik kredit riski məlumatlarımızın nəticəsini göstərir. Sadəlik məqsədilə 11 dəyişəndən istifadə etdik. Bu, simulyasiya edilmiş məlumatlara bənzər bir yola gətirib çıxarır.

Lasso Reqresiyası

Statistikada və maşın öyrənməsində, Lasso (ən az mütləq büzülmə və seçim operatoru) nəticələnən statistik modelin proqnozlaşdırılmasının düzgünlüyünü və şərh edilə bilənliyini artırmaq üçün həm dəyişən seçimi, həm də nizamlanmanı həyata keçirən reqressiya analizi üsuludur. Lasso metodu xətti modelin əmsallarının seyrək olduğunu, yəni onlardan bir neçəsinin sıfırdan fərqli olduğunu nəzərdə tutur. Əvvəlcə, geofizikada daha sonra bu termini ortaya qoyan Robert Tibşirani tərəfindən təqdim edilmişdir. Ric reqressiyasının əksinə olaraq, biz burada büzülmə üçün parametrlərin mütləq qiymətlərindən istifadə edirik.

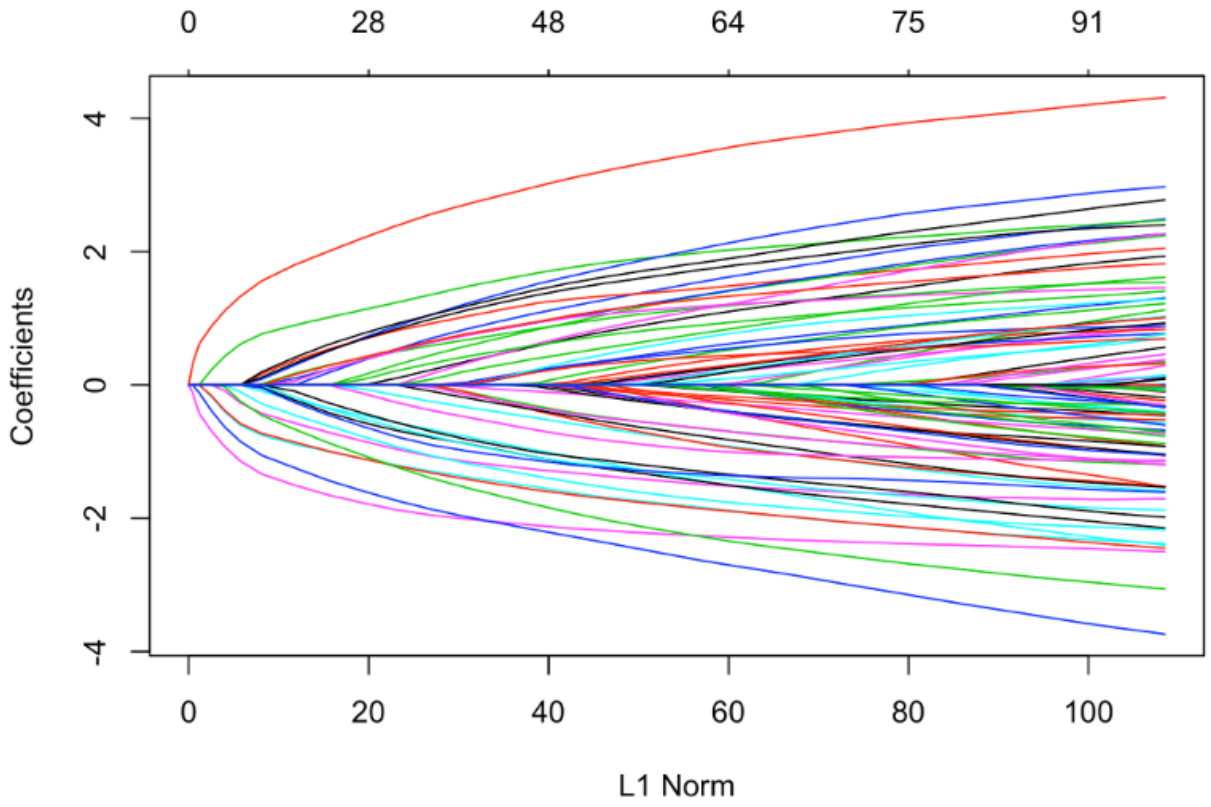
$$\min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{N} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}$$

Nəticədə biz bəzi parametrləri sıfıra endirmiş oluruq. Bu, Lasso-nun maraqlı xüsusiyyətləridir, çünki o, vacib

dəyişənləri (sıfıra endirilməmiş) seçməyə kömək edir. Bu, bizə əhəmiyyət testi üçün p dəyərindən başqa müasir yanaşma təqdim edir. Lambda parametrindən asılı olaraq modeldə və Lassoda dəyişənlərin böyük hissəsindən istifadə edə bilərik, təbii ki, bunların hamısı bizə hansı dəyişənləri daha çox vurğulamalı olduğumuzu söyləyəcək. İndi əmsal yollarına baxaq.

Qrafik 4. Simulyasiya olunmuş Lasso üçün əmsal Yolları

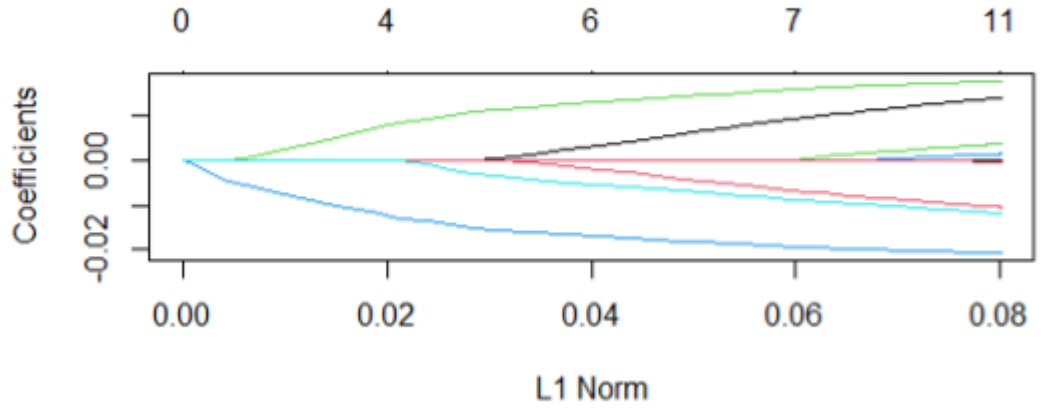
Yuxarıda biz simulyasiya edilmiş 100 dəyişəndən əmsal yollarını görürük. Gördüyünüz kimi, o, Ric əmsallarının yollarından çox fərqlidir. Lambdaya görə, model əvvəlcə 28 dəyişəni seçir, sonra isə model 48 dəyişəni vacib və sıfıra enməmək üçün seçir. Əgər biz p dəyərləri ilə dəyişən seçimi etsəydik, 100-dən 28 dəyişəni olan bir model əldə etmək üçün qeyri-adi dərəcədə vaxt və səy sərf edərdik. İndi biz bunu bir proqram hissəsiylə sadə şəkildə edirik və bu, Lasso modelləşdirməsinin effektivliyidir.



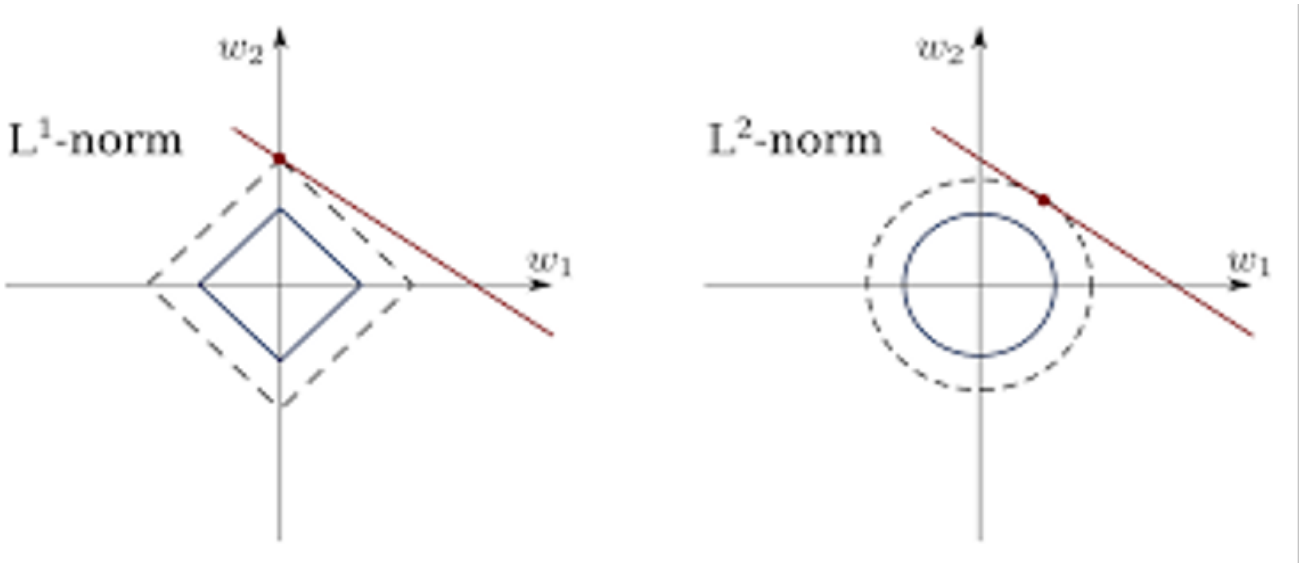
Qrafik 5. Kredit Riski Datasının Lasso əmsal Yolları

Yuxarıdakı real dünya kredit riski məlumatlarımızın

nəticəsidir. Qrafikdən aydın olur ki, Lasso əvvəlcə 4 dəyişən seçir, sonra isə lambda tənzimləmə parametrini dəyişdikcə dəyişənlərin sayı artıb. Model simulyasiya edilmiş məlumatlarla da uyğun gəlir.



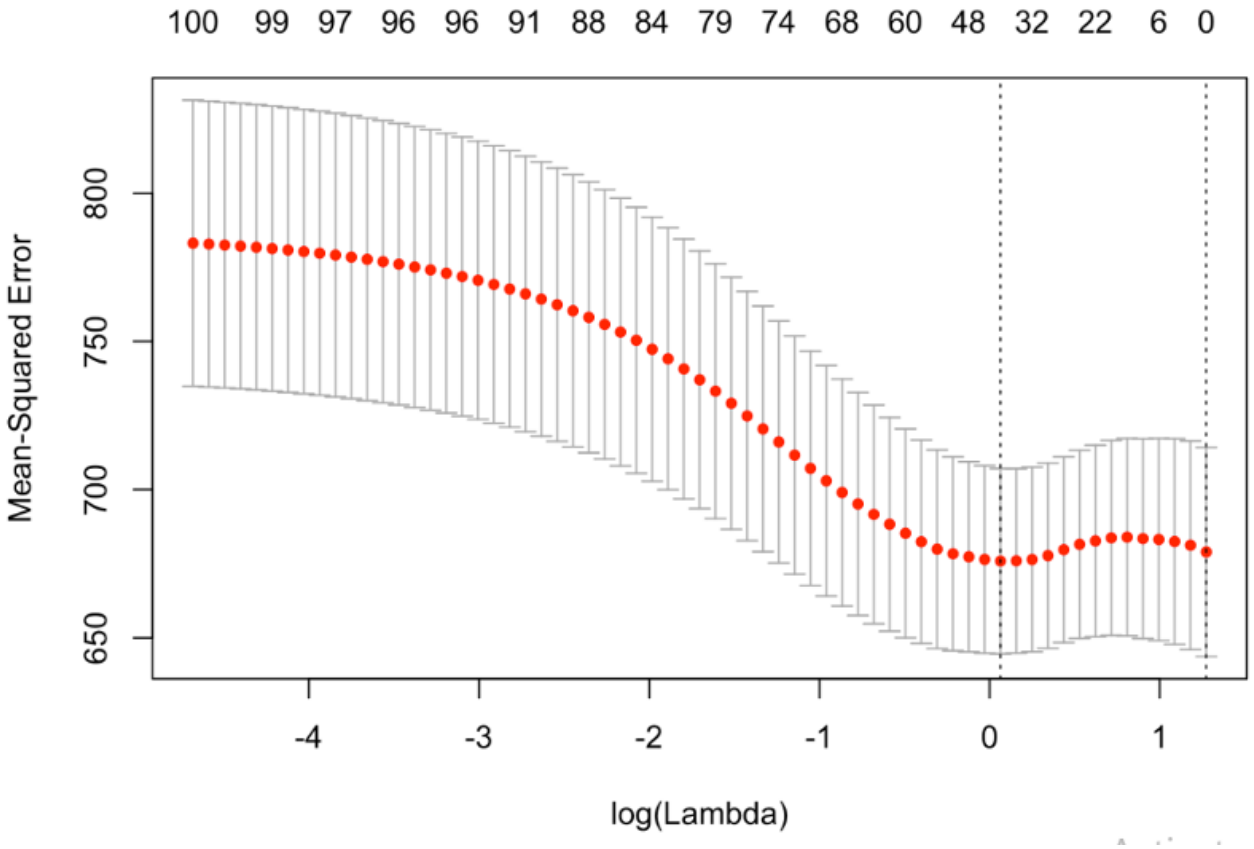
Qrafik 6. Reqularizasiya Mexanizmləri



Yuxarıda Lasso və Ric reqressiyaları arasındakı fərqi göstərən qrafiklər var. Birincisi- Lasso (L1 norması), ikincisi -Ric (l2 norması). Qrafiklərdən aydın olur ki, niyə Lasso modeldə bir parametri sıfır seçir, Ric isə hər ikisini seçir.

Aşağıda lambda başına orta kvadrat xətanı göstərən Lasso modeli üçün çıxış verilmişdir. Lambda vasitəsilə parametrlərin necə azaldığını və orta kvadrat xətanın necə minimuma endirildiyini görmək əladır.

Qrafik 7. Xəta Minimalizasiyası və Dəyişən Seçimi



Yekun olaraq, Lasso və Ric reqressiya üsullarının tədqiqi onların statistik modelləşdirmə və maşın öyrənməsi sahəsində güclü alətlər kimi əhəmiyyətini işıqlandırır. Bu yazı boyu biz Lasso və Ric reqressiyasının nəzəri əsaslarını, metodologiyalarını və praktiki tətbiqlərini araşdırdıq, modelin möhkəmliyini, şərh oluna bilənliyini və proqnozlaşdırıcı performansını artırmaqda onların rollarına işıq saldıq.

Lasso və Ric reqressiyası reqressiya təhlilinə xas olan problemlərə, o cümlədən həddən artıq uyğunlaşma, multikollinearlıq və yüksək ölçülü məlumatlar üçün nəfis həllər təklif edir. Reqressiya əmsallarının böyüklüyünü cəzalandıran nizamlanma cəzalarını tətbiq etməklə, bu üsullar qeyri-müəyyənlik və variasiya arasında zərif tarazlıq yaradır və şərh edilə bilənliyi qoruyarkən görünməyən məlumatlara yaxşı ümumiləşən modellərə gətirib çıxarır.

Praktik nöqteyi-nəzərdən, Lasso və Ric reqressiyası xüsusiyyət

seçimi, dəyişən əhəmiyyətin qiymətləndirilməsi və model sabitliyi üçün dəyərli üstünlüklər təqdim edir. Lasso reqressiyasının L1 cəzası müəyyən əmsalları sıfıra endirərək, avtomatik xüsusiyyət seçimini asanlaşdıraraq və modelin şərh edilməsini gücləndirərək seyrəkliyi artırır. Eyni zamanda, Ric reqressiyasının L2 cəzası multikollinearlığın təsirlərini azaldır, qiymətləndirmə prosesini sabitləşdirir və korrelyasiya edilmiş proqnozlaşdırıcı dəyişən ssenarilərində model performansını yaxşılaşdırır.

Müxtəlif sahələr üzrə Lasso və Ric reqressiya üsullarının geniş tətbiqi onların real dünya reqressiya problemlərinin həllində çox yönlü və effektivliyini vurğulayır. Maliyyə və iqtisadiyyatdan səhiyyə və mühəndisliyə qədər bu üsullar məlumatlara əsaslanan anlayışların və proqnozlaşdırıcı modelləşdirmənin əsas olduğu müxtəlif sahələrdə öz tətbiqini tapır.

Gələcəyə baxaraq, müntəzəmləşdirmə üsullarında davamlı tədqiqat və yenilik Lasso və Ric reqressiyasının imkanlarını daha da artırmağı vəd edir. Hesablama alqoritmləri, optimallaşdırma metodları və modellərin şərh edilməsi alətlərindəki irəliləyişlər bu texnikaların tətbiqini və əlçatanlığını genişləndirməyə davam edəcək, tədqiqatçılara, analitiklərə və praktikantlara mürəkkəb verilənlər toplularından mənalı fikirlər çıxarmaq imkanı verəcək.

Xülasə, Lasso və Ric reqressiyasının tədqiqi müasir məlumat elminin mühüm hissəsidir, çünki ikisi möhkəm, şərh edilə bilən və proqnozlaşdırıla bilən reqressiya modellərinin qurulması üçün mükəmməl çərçivə təklif edir. Bu üsulların prinsiplərini və tətbiqlərini başa düşməklə, biz onların gücündən müasir məlumat təhlilinin mürəkkəbliklərini həll etmək və biliklərin kəşfi, o cümlədən qərar qəbulu üçün yeni imkanlar açmaq məqsədilə istifadə edə bilərik.

İstinadlar:

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.2307/1267351

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. http://www.jstor.org/stable/2346178

Kredit riskinin modelləşdirilməsi: PD modelləşdirməsində daralma üsulları və LASSO seçimi (1-ci hissə)