KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

دراسة اهم العوامل المؤثرة على الجلطة الدماغية باستخدام بعض طرائق التنظيم الجزائية

أ.د. احمد نعيم فليح
الباحث حيدر ناظم محمد علي
جامعة القادسية / كلية الادارة والاقتصاد

DOI: https://doi.org/10.36322/jksc.176(G).20052

الملخص

تم في هذا البحث دراسة مفهوم أسلوب اختيار المتغيرات من خلال توظيف طرائق الجزاء penalty (method) في نموذج الانحدار الخطي المتعدد . حيث تم تناول أربعة طرائق تنظيم (جزاء) وهذه الطرائق هي (طريقة ridge ، طريقة lasso ، طريقة adaptive lasso ، طريقة elastic net). تم أيضا توضيح كل طريقة جزاء من خلال دراسة صيغتها الرياضية وكذلك اهم الخصائص والعيوب لكل طريقة .

وبهدف دراسة هذه الطرائق تم أجراء تجربة محاكاة واحدة في هذا البحث تم افتراض وجود متجه حقيقي للمعالم المراد تقدير قيمها بهدف قياس دقة عمل كل طريقة ، أجريت هذه التجربة بافتراض توليد حجوم عينات مختلفة ، حيث تم أولا تحويل مشاهدات المتغيرات التوضيحية إلى القيم المعيارية لفك الارتباطات الموجودة بينها إضافة إلى ذلك تم أجراء تحويل المشاهدات متغير الاستجابة إلى قيم مركزية (centering) ، أي ان الدراسة افترضه ان هناك قيم مختلفة من الارتباطات بين المتغيرات التوضيحية. إشارة النتائج في تجارب المحاكاة إلى ان طريقة elastic net قد أظهرت اقل قيمة لمعيار جودة تقدير

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

المعالم المسمى معيار وسط متوسط الفروق المطلقة ، وبالتالي يمكن اعتبار هذه الطريقة هي افضل الطرائق من ناحية دقة وتفسير النموذج .

ولبيان كفاءة عمل طرائق الجزاء المذكورة أعلاه تم تحليل لبيانات حقيقية قمت بجمعها من مستشفى الناصرية التعليمي لتمثل عينة من المرضى المصابين بالجلطة الدماغية على اعتبار ان متغير ان متغير الاستجابة في هذه الدراسة يمثل حجم الجلطة الدماغية ومجموعة من (١٧) متغير توضيحي ، وتمثيل العلاقة بين متغير حجم الجلطة والمتغيرات التوضيحية بنموذج انحدار خطي متعدد من اجل الوقوف على اهم العوامل أو المتغيرات التي تؤثر على حجم الجلطة الدماغية . كذلك أظهرت النتائج ان طريقة elastic net هي افضل طريقة جزاء سوآءا على مستوى اختيار المتغيرات أو على مستوى القدرة التفسيرية للنموذج .

الكلمات المفتاحية : الجلطة الدماغية ، طرائق التنظيم الجزائية ، الانحدار الخطي

Studying the most important factors affecting stroke by using some penal regulation methods

Prof. Dr. Ahmad Naeem Flaih
Researcher Hayder Nadhim Mohamed Ali
University of Al-Qadisiyah / College of Administration and Economics

Abstract

In this paper the concept of variable selection method was studied by employing the penalty method in the multiple linear regression model. Four organizing

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

methods (parts) were dealt with, and these methods are (the ridge method, the lasso method, the adaptive lasso method, and the elastic net method). Each penalty method was also clarified by studying its mathematical formula, as well as the most important characteristics and defects of each method.

In order to study these methods, three simulated experiments were conducted. In each experiment, it was assumed that there is a real vector for the parameters whose values are to be estimated in order to measure the accuracy of the work of each method. Among them, in addition to that, the conversion of the observations of the response variable into central values was performed, that is, the study assumed that there are different values of the correlations between the explanatory variables. The results of the simulation experiments indicated that the elastic net method showed the lowest value for the standard of feature estimation quality called the standard of mean absolute differences, and therefore this method can be considered the best method in terms of accuracy and interpretation of the model.

In order to demonstrate the efficiency of the work of the penalty methods mentioned above, real data collected by the researcher from Al-Nasiriyah Teaching Hospital were analyzed to represent a sample of patients with stroke,

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

given that the response variable in this study represents the size of the stroke and a group of (17) explanatory variables, and the representation of the relationship between Stroke size variable and explanatory variables using a multiple linear regression model in order to identify the most important factors or variables that affect the size of stroke. The results also showed that the elastic net method is the best penalty method, whether at the level of choosing the variables or at the level of the explanatory ability of the model.

Keyword : Linear Regression , Penalized Regression Methods, Stroke

المقدمة

تعتبر صياغة النماذج الإحصائية من اهم وسائل الباحثين في مجال تحليل البيانات لمشاكل محددة . حيث عادة ما يعتبر نموذج الانحدار عن العلاقة بين متغير الاستجابة Y والمتغيرات التفسيرية X . وبهذا يمكن القول أن تحليل الانحدار هو طريقة بسيطة للتحقق من وجود علاقة دالية بين المتغيرات . عادة ما يفترض أن متغير الاستجابة يخضع الى الخطأ ($\in$) وأن الخطأ و X يكونان مستقلان . وفي سياق رسالتنا سنفترض أن الأخطاء هي متغيرات عشوائية و بأن $E(\in) = 0$ ، ومن هنا يمكن القول ان دالة الانحدار تحاول أيجاد تقدير متوسط متغير الاستجابة $E(Y|X)$ من خلال استخدام المعلومات المتوفرة في المتغيرات التفسيرية X بافتراض خطية العلاقة بين Y و X .أن القيمة المحسوبة العددية (estimate) للمعلمة المقدرة تعتمد في نوعيتها على بعض الجوانب الحاسمة ، منها دقة التنبؤ حيث ما يعتمد هذا المفهوم كل عملية المتبادلة بين تباين المعلمة المقدرة و ما بين تحيز المعلمة المقدرة ، والجانب الحاسم الاخر هو القدرة

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

التفسيرية لنموذج الانحدار . تغير طريقة المربعات الصغرى(Ordinary least squares) او مختصراً (OLS) من أشهر الطرائق المستخدمة في أيجاد قيم مقدرات معالم نموذج الانحدار الخطي وهي طريقة بسيطة تعتمد على مبدء تقليل مجموع مربعات دالة الخسارة أو تقليل مجموع مربعات البواقي (residual sum of squares) مختصراً تكتب (RSS). من المعلوم أن طريقة OLS توفر لنا مقدرات تميز خاصية BLUE وتعني أن هذه الطريقة تعطي تقديرات لمعالم نموذج الانحدار تتميز بأنها تقديرات غير متحيزة و تمثل أفضل خط انحدار للبيانات المدروسة وهذا لن يتم الا بتوفر شروط Gauss (Markov) والتي تنص على توفر الشروط الأتية (الخطية في المعالم ، عشوائية المتغيرات التفسيرية،عدم وجود مشكلة التعدد الخطي ، ان $E(x_i|\in_i) = 0$ ، ثبات تباين الأخطاء) .

و بهذا فأن طريقة OLS تتميز بأنها غير متحيزة و تمتلك أقل تباين و بالتالي فأنها تعطى أفضل دقة للتنبؤ تعمم متغير الاستجابة Y . ومن الجدير بالذكر و عندما تكون كمية المعلومات المتوفرة من البيانات أو (حجم العينة) أكبر من عدد المتغيرات التفسيرية فأن تباين المقدرات ستقل قيمته مما يعكس جودة دقة التنبؤ . كذلك فأن خرق بعض شرط (Gauss Markov) أو عندما تكون كمية المعلومات المتوفرة من العينة قليلة (أي عدد المتغيرات التفسيرية قريب من حجم العينة) فإن مقدرات طريقة OLS تتبع دقة تفسيرية غير مفيدة (poor) ، إضافة إلى ذلك فإن طريقة OLS لا يمكن استخدامها عندما يكون عدد المتغيرات التفسيرية (k) اكبر من حجم العينة (n) .

فمثلا خرق شرط التعدد الخطي (collinearity) يؤدي إلى تضخم تباين المقدرات مما ينعكس سلبا على دقة تنبؤ نموذج الانحدار ، و بالتالي يمكن حذف بعض المتغيرات التفسيرية التي ليس لها تأثير على تغيرات متغير الاستجابة حيث نلاحظ أن بعض الدراسات قد تحدد المتغيرات التفسيرية الداخلة في تحليل

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

نموذج الانحدار بدون أي أساس نظري (أي أنها غير محددة مسبقاً) لذلك يمكن في هذه الحالة الاعتماد على أسلوب اختيار المتغيرات (Variable selection procedure) في اختيار المتغيرات الأكثر تأثيرا . وبالتالي فأنه اختيار المتغيرات هو موضوع لا ينفصل عن اختيار النماذج و بهذا يمكن القول أن خرق شروط نظرية (Gauss Markov) أي أن عدد المتغيرات التفسيرية أكبر من عدد المشاهدات (n) يؤدي الى أن تطبق طريقة OLS لنمذجة خط الانحدار تكون غير مناسبة . وبالنتيجة هذا صفن الباحثين للتغلب يهدف هذا البحث الى :

تحديد المتغيرات التفسيرية المهمة أي التي لها تأثير على متغير الاستجابة و بالتالي تحديد أفضل نموذج انحدار يمثل العلاقة ذات الصلة من خلال تسليط الضوء على أهم المتغيرات التفسيرية التي تفسر أسباب الجلطة الدماغية .

و دراسة أثر طرائق التنظيم الجزائية في الحصول على مقدرات ينتج عنها نموذج انحدار ذو قدرة تفسيرية عالية .

و أجراء دراسة مقارنة بين الطرائق الجزائية لأختيار النماذج مستخدمين طريقة لاسو ، و لاسو التكيفية ، وridge ، و elastic net في كل من جانب المحاكاة وجانب البيانات الحقيقية .

على هذه الصعوبات من خلال اقتراح طرائق بديلة أو أجراء تعديلات على طريقة (OLS) . حيث تعتبر طرائق التنظيم (Regularization) من أشهر الطرائق للتغلب على صعوبات استخدام طريقة OLS . أن طرائق التنظيم عادة ما تسمى بطرائق الجزاء (penalty methods) ، حيث تحاول هذه الطرائق على أيجاد مقدرات لمعالم نموذج الانحدار تتميز بأنها تعطي مقدرات ذات تباينات قليلة بوجود تحيز في قيمتها المقدرة .

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

١- الانحدار الخطي (Linear Regression)

تعتبر منهجية الانحدار طريقة لدراسة العلاقة بين متغير الاستجابة (Y) ومجموعة من المتغيرات التفسيرية (X) ، وهذا يعني انه كمن وصف نموذج الانحدار الخطي المتعدد بالمعادلة الأتية وبصيغة المصفوفات :

$$Y = X\beta + \in \qquad \ldots (1)$$

أي أن معادلة الانحدار تهدف الى تقدير متوسط متغير الاستجابة باستخدام المعلومات المتوفرة في المتغيرات التفسيرية X . حيث يفترض النموذج (1) بوجود خطية العلاقة بين Y , X وبين β , Y . ومن الواضح أن قيمة المعالم (β) هي غير معلوم القيم و يجب أن يتم تقدير قيم معالمه . في عام ١٩٧٠ قدم الباحثان (Kennard و Hoerl) طريقة الجزاء أو التنظيم المسماة طريقة (Ridge) وهي طريقة ناقشت مشكلة التعدد الخطي Multicollinearily و معالجتها من خلال أضافة دالة جزاء الى مجموع مربعات الخطأ . و تستعمل كذلك هذه الطريقة عندما يكون عدد المتغيرات التوضيحية قريبة جدا الى حجم العينة . أن مقدرات هذه الطريقة تكون متجزة لكنها ذات تباينات قليلة مما ينعكس على دقة التنبؤ ، أي أنها تعمل على جعل التباينات المقدرات مستقرة . كذلك فأن دالة الجزاء يتم من خلالها التحكم بمقدار قيمة المقدرات و تقليصها وفق الأتي :

$$\hat{\beta}_{ridge} = argmin\left[RSS + \lambda \sum_{j=1}^{p} \beta_j^2\right]$$

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

حيث أن معلمة الجزاء أو الانكماش shrinkage هي $\lambda \geq 0$ و تدعى أيضا معلمة الضبط tuning . أن طريقة (ridge) لا تقلص قيم معالم نموذج الانحدار و تكون مساوية للصفر أي أنه لا يمكن اعتبار طريقة (ridge) كإحدى طرائق اختيار المتغيرات .

في عام ١٩٩٦ قدم الباحث (Tibshirani) بحثاً ناقش فيه طريقة جديدة مقترحة في موضوع اختيار المتغيرات و هذه الطريقة تدعى طريقة (Lasso) ، هذه الطريقة توفر أسلوب اختيار المتغيرات مما يعني أنتاج نموذج انحدار ذو قدرة تفسيرية عالية مع أنتاج مقدرات بتباينات قليلة مما يجعل النموذج ذو قدرة تنبؤيه عالية . أساس عمل هذه الطريقة هو تقليص معالم المتغيرات المهمة التي لها تأثير على متغير الاستجابة و بالتالي التحكم بتباين المقدرات ، أضافة إلى ذلك فأنها تجعل قيم المقدرات للمتغيرات التي ليس لها تأثير على متغير الاستجابة مساوية للصفر مما تتبع عنه نموذج ذو قدرة تفسيرية عالية ، أن مقدر طريقة يدعى هو :

$$\hat{\beta}_{Lasso} = argmin\left[RSS + \lambda \sum_{j=1}^{p} |\beta_j|\right].$$

أضافة الى ذلك فأن مقدرات طريقة (lasso) لا تتصف بكونها oracle (خاصية الاتساق consistent وخاصية asymptotic normal) . ومن الجدير بالذكر أن دالة الجزاء في طريقة (Lasso) هي دالة غير قابلة للاشتقاق عند الصفر ، وكذلك فأنه لا توجد صيغة تقريبية لمقدرة لاسو وكذلك للخطأ المعياري لمقدر لاسو . كذلك فأن قيمة مقدر لاسو يعتمد بشكل كبير على اختيار قيمة المعلمة λ والتي يمكن حسابها من خلال طريقة Cross-Validation (CV) .

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

٢- طرائق انحدار الجزاء (Penalized Regression Methods)

بينا في المبحث السابق أن طريقة المربعات الصغرى تعمل و تعطي حلاً وحيداً عندما يكون عدد المتغيرات التفسيرية أقل من حجم العينة هذا يعني أن طريقة المربعات الصغرى تعتمد بشكل كبير على حجم العينة . في كثير من الأحيان وفي الواقع العملي عادة ما يواجه محلل البيانات أن هناك عدد كبير من المتغيرات التفسيرية يفوق حجم العينة وبالتالي تكون المصفوفة $\acute{x}x$ مصفوفة singular وبالتالي لا يمكن أيجاد حل وحيد لمعالم النموذج ، أو قد يكون هناك ارتباطات بين المتغيرات التفسيرية مما يسبب بمشكلة التعدد الخطي Multicollinearily . وكل هذا يسبب بالنتيجة ضعف في القوة التفسيرية للنموذج وضعف في دقة التنبؤ، وبالتالي مثل هذه المشاكل حفزت الباحثين للتفتيش عن طرائق لتقدير معالم نموذج الانحدار يمكنها التغلب على عيوب طريقة OLS . تعطى للمحدد (Flexeder , 2010) للمزيد من هذه المعلومات . حيث يشير هذا الباحث إلى أنه حتى لو عملت طريقة OLS بوجود هذه المشاكل فأنها سوف تنتج مقدرات ذات تحيز قليل وتباينات عالية . فاذا كان الهدف من بناء النموذج هو الدقة التنبؤ فـأن ما يسمى طرائق بطرائق التنظيم (regularization modalities) تهتم بمعالجة هذه المشاكل ، حيث أنها توفر صفقة عادلة (trade off) ما بين التحيز والتباين . طرائق التنظيم تسمى أيضاً بطرائق الجزاء، حيث ان المقدر وفق هذه الطرائق هو الذي يجعل جزاء مجموع مربعات الخطأ أقل ما يمكن ، أي ان :

$$\hat{\beta}_{penalizd} = argmin_{\beta}[RSS(\hat{a}) + \lambda(penalty\ function)] \quad \ldots. (2)$$

حيث أن $\lambda \geq 0$ تسمى معلمة الضبط (tuning parameter) أو معلمة الانكماش (shrinkage) أو معلمة التنظيم (regularization) . فاذا كانت قيمتها مساوية للصفر فأن مقدار OLS يتم الحصول عليه . حيث تعمل هذه المعلمة على ضبط كمية الانكماش بحجم المعلمة (β) ، أي أنها تكبر و تصغر

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

المعلمة (β) . فعندما $\lambda \to \infty$ فهذا يؤدي الى أن تكون (β) قليلة جدا و العكس صحيح . أي أنها تضبط تأثير دالة الجزاء على قيمة المعلمة (β) . وتوجد عدة دوال جزاء فمثلا دالة جزاء ridge ، دالة جزاء lasso ، ودالة جزاء elastic net حيث سيتم اعتماد هذه الطرائق في هذه الرسالة . في طرائق التنظيم يعاني الباحث عادة من مشكلة حساب قيمة معلمة الانكماش أو التقليص أو الضبط (λ) ، حيث يعتبر هذا الموضوع مهماً جداً و يجب الاهتمام به . حيث توجد هناك عدة طرائق لهذا الغرض ، منها الطرائق الكلاسيكية مثل ، AIC ، BIC و طريقة generalize at cross validation (G.C.V) ، وطريقة k-fold cross validation

A: طريقة الجزاء (Ridge)

تم اقتراح هذه الطريقة من قبل الباحث (Hoerl , kennard) في عام ١٩٧٠ وهي أقدم طريقة جزاء وتعرف أيضا بطريقة الانكماش (shrinkage method) . حيث تفرض هذه الطريقة قيداً على حجم المعالم و بالتالي مما يؤثر على تقليص أو انكماش قيم المعالم لتكون قريبة من الصفر لكنها لا يمكن ان تكون مساوية للصفر، كذلك فأنها لا تعتبر من طرائق اختبار المتغيرات . وتتعامل طريقة (Ridge) في حالة كون أن هناك مشكلة تعدد خطي تعاني منها المتغيرات التفسيرية ، وكذلك تتعامل في حالة وجود عدد كبير من المتغيرات التفسيرية ، الأمر الذي يجعل مصفوفة $(\acute{X}X)$ مصفوفة singular وبالتالي لا يمكن أيجاد معكوسها . وبالتالي فأن هذه الطريقة تعمل على تقليص تباين المقدرات إلى أقل ما يمكن بوجود بعض التحيز ومن الجدير بالذكر فأنه يجب تحويل المتغيرات التفسيرية إلى قيم معيارية قبل البدء بتقدير المعالم بطريقة Ridge

$$\hat{\beta}_{ridge} = argmin_{\beta} \|y - x\beta\|^2 + \lambda \sum_{j=1}^{P} \hat{\beta}_j^2 \qquad \dots (3)$$

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

B: طريقة الجزاء (Lasso)

تعتبر طريقة لاسو Lasso من اكثر الطرائق الجزائية استعمالاً كأجراء لاختبار المتغيرات Variable Selection وقام الباحث في عام ١٩٩٦ Tibshirani باقتراح هذه الطريقة حيث تختلف هذه الطريقة باستعمالها L_1-nom في دالة الجزاء بدلاً L_2-norm في طريقة الجزاء ridge . وتعتبر طريقة التنظيم لاسو كطريقة تقدير للمعالم وأسلوب لاختبار المتغيرات التي لها تأثير على متغير الاستجابة وإهمال المتغيرات التي ليس لها تأثير على متغير الاستجابة ، مما ينتج عنها نموذج انحدار مقدر ذو قدرة تفسيرية عالية . وكما في طريقة الجزاء ridge فأن دالة جزاء lasso لا تحتوي على المعلمة حيث يتم ذلك من خلال عمل centering لكل المتغيرات التفسيرية وتغيير الاستجابة . ان مقدر لاسو يعمل على تقليل مجموع مربعات الخطاء RSS .

$$\hat{\beta}_{Lasso} = argmin_{\beta}\|Y - X\beta\|_2^2 \quad \ldots(4)$$

وفقا للتعبير $\|\beta\|_1 \leq t \; ; \; t \geq 0$ حيث ان $\|\beta\|_1 = \sum_{j=1}^{P} \beta_j$ ويدعى L_1-norm أوبصيغة أخرى فأن مشكلة التصغير (٤) (Minimization problem) يمكن إعادة كتابتها بالصيغة الأتية :

$$\hat{\beta}_{\text{Lasso}} = argmin_{\beta}\|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \quad \ldots(5)$$

حيث ان $\lambda \geq 0$ ، حيث عندما $\lambda = 0$ فأن المقدر يصبح مقدر طريقة OLS .

C: طريقة جزاء الشبكة المرنة Elastic Net

قدم الباحثان (Zou و Hastie) في عام ٢٠٠٥ طريقة جزاء جديدة تدعى الشبكة المرنة (elastic net) وهي احدى طرائق التنظيم تم اقترحها كبديل لطريقة lassoعند تحليل الانحدار للبيانات الخاصة بدراسة الجينات genes الوراثية . ان طريقة الشبكة المرنة تعمل كأسلوب لاختيار المتغيرات

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

(Variable Selecting) وفي نفس الوقت تعمل على تقليص المعالم الأخرى . كذلك فأنها تعمل على اختبار مجموعات من المتغيرات التفسيرية المرتبطة وبالتالي يمكن ان تتغلب على مشكلة ان عند المتغيرات التوضيحية اكبر من عند المشاهدات . وان الية عمل طريقة الشبكة المرنة هو بالأساس يعتمد دمج دالة الجزاء ridge ودالة جزاء lasso ، وبالتالي فأن مقدر الشبكة المرنة يمكن ترتيبه كالاتي :

$$\hat{\beta}_{elastic\ net} = (1+\lambda_2) argmin_{\beta}[\|Y - X\beta\|_2^2 + \lambda_2\|\beta\|_2^2 + \lambda_1\|\beta_1\|]\} \quad ..(٦)$$

حيث ان $\lambda_1 \geq ٠$ و $\lambda_2 \geq ٠$ هي معالم دوال التنظيم او التقليص ، حيث ان λ_1 تتحكم بكمية التقليص واختبار المتغيرات و λ_2 تتحكم بكمية grouping

D: طريقة الجزاء لاسو التكيفية Adaptive lasso

تقوم طريقة لاسو على جعل عدد قليل من المعالم مساوياً للصفر والمعالم الأخرى يتم تقليص قيمتها عند معدل ثابت بغض النظر عن عددها ، وبالتالي فأن عدد كبير من المعالم سيتم تقليص قيمتها مما يسبب بزيادة تحيز مقدرات طريقة لاسو . ان مشكلة طريقة لاسو هو اعتمادها على استخدام معلمة ضبط (أو تقليص أو تنظيم) واحدة لكل من أسلوب اختيار المتغيرات وتقدير المعالم .

ان اختيار المعلمة λ بطريقة Cross – validation أو طريقة معايير المعلومات information Criteria)) ينتج عنه نموذج مثالي للتنبؤ .

ان قيمة المعلمة λ غالباً ما تكون صغيرة في هذه الحالة وبالتالي يتم استبعاد عدد قليل من المتغيرات من النموذج . اما اذا كانت قيمة λ كبيرة فهذا يعني جعل عدد كبير من معالم المتغيرات قيمتها مساوية للصفر وبالتالي استبعادها من النموذج . كل هذا أدى إلى اقتراح ما يسمى بطرائق لاسو التكيفية التي تستخدم اكثر معلمة تقليص واحدة .

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

حيث اقترح الباحث (Zou) في عام ٢٠٠٦ طريقة الجزاء المسماة لاسو التكيفية التي تستخدم وزن وزن موجب القيمة للسيطرة او لضبط التحيز من خلال تقليص قيم المعالم بواسطة كميات مختلفة . ان مقدر طريقة لاسو التكيفية معطى بالاتي :

$$\hat{\beta}_{adaptiv.Lasso} = argmin_{\beta}\|y - x\beta\|^2 + \lambda \sum_{j=1}^{P} \hat{w}_j \left|\beta_j\right| \ldots(7)$$

حيث ان $\lambda \geq 0$ هي معلمة الانكماش او التقليص وان $W_j > 0$ هي الاوزان .

وقد اقترح (Zou) عام ٢٠٠٦ استخدام القيمة التقديرية الآتية :

$$\hat{w}_j = \frac{1}{\left|\hat{\beta}_j\right|^{\alpha}}$$

حيث ان $\alpha > 0$ وان $\hat{\beta}_j$ هي القيمة الابتدائية او الاساسية (initial) المقدرة للمعلمة β ، عادة فيتم اختيار $\hat{\beta}_j$ لتكون قيمة المعلمة من طريقة OLS او طريقة lasso.

الجانب التطبيقي

يتضمن هذا القسم تطبيقا عمليا لطراق التقدير(ridge, lasso, adaptive lasso, elastic net) على بيانات من الواقع الحقيقي عن السكتة الدماغية (التخثر) المسجلة للمرضى حيث تم سحب عينة عشوائية من المرضى لسنة ٢٠٢٢ من شعبة الاحصاء في مستشفى الناصرية التعليمي التابع الى دائرة صحة ذي قار في محافظة ذي قار من تاريخ ٢٠٢٢/١٠/٢٦. وتم تقدير نموذج الانحدار الخطي المتعدد لتمثيل العلاقة بين حجم الجلطة الدماغية كمتغير استجابة ومجموعة من المتغيرات التفسيرية باستخدام

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

طرائق الجزاء اعلاه وبالتالي تحديد افضل نموذج يمثل المتغيرات التي لها تأثير مباشر على متغير الاستجابة.

٣: الجلطة الدماغية (Stroke)

وتعد الجلطة الدماغية على انها من الأمراض الشائعة في وقتنا الحالي والسابق وتكون من ضمن اخطر الأمراض على الأنسان ويجب معالجتها مبكرا حسب رأي أطباء الجملة العصبية ليتجنب المريض الإعاقة الذهنية او الإعاقة الحركية ويمكن تعريفها هي عبارة عن اضطراب يصيب خلايا المخ ناتج عن خلل في توصيل الدم إلى المخ بأي شكل من الأشكال وتسمى أيضا بالسكتة الدماغية .

وتنقسم السكتة الدماغية الى نوعان رئيسيان:

١- السكتة الدماغية الأقفارية (Ischemic Stroke) : وهي نقص في كمية الدم والأوكسجين المنقول الى الدماغ بسبب الترسبات الدهنية او تخثر الدم في الشريان وهو نوع من انواع السكتات الدماغية الأكثر شيوعا بنسبه ٨٠٪ . وينقسم هذا النوع إلى قسمين

- السكتة الدماغية الاختارية (Thrombotic Stroke)
- السكتة الدماغية الصمية (Stroke embolic)

٢- السكتة الدماغية النزفية (Intracerebral Hemorrhage) : وهذا النوع من السكتات يكون بشكل نزف داخل الدماغ بسبب تسريب او تمزيق او انسداد احد الأوعية الدموية في الدماغ . وينقسم هذا النوع الى قسمين ايضا

- نزيف داخل الدماغ
- النزيف في المنطقة تحت العنكبوتية

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

وهناك العديد من العوامل التي تزيد من خطر الاصابة بالسكتة الدماغية ومنها

- عوامل خطر بعض الامراض

ارتفاع ضغط الدم و ارتفاع داء السكري و ارتفاع نسبة الكولسترول و ارتفاع هرمون الدي دايمر و حبوب منع الحمل يزيد من خطر الاصابة و عدوى كوفيد ١٩ و امراض القلب الوعائية مثل (الرجفان الاذيني)

- عوامل الخطورة اثناء حياتك

السمنة او زيادة الوزن و قلة النشاط الرياضي و تناول الكحول بكميات كبيرة و تدخين السجائر و تعاطي المخدرات و الجنس تكون نسبة اصابة الرجال اكثر من النساء ويزداد هذا الخطر بعد ٥٥ سنة

الجدول في ادناه يوضح القيم التقديرية لمعالم النموذج الخطي المتعدد.

جدول رقم (١): قيم المعالم المقدرة لمتغيرات النموذج

Name of variables	Symbol variables	Ridge Reg	Lasso Reg	Adaptive lasso Reg	Elastic-Net-Reg
العمر	x_1	-0.0173	0.0000	0.0000	0.0000
الجنس	x_2	0.0659	0.0000	0.0000	0.00203
الوزن	x_3	-0.0292	0.0000	0.0000	0.0000
التدخين	x_4	0.0142	0.0000	0.0000	0.0000
الضغط	x_5	-0.0667	-0.0097	-0.13077	-0.02423
داء السكري	x_6	0.0553	0.0000	0.0000	0.01400
دي دايمر	x_7	0.2175	0.1496	0.6326	0.16713
كوليسترول	x_8	-0.0869	-0.0529	-0.58770	-0.06560
الرجفان الأذيني	x_9	0.0603	0.0345	0.55132	0.04142
عدوى كوفيد	x_{10}	0.0457	0.0000	0.0000	0.005087
مدة رقود المريض بالأيام	x_{11}	-0.0355	0.0000	0.0000	0.0000
سكن المريض	x_{12}	-0.0168	0.0000	0.0000	0.0000
محل الاقامة	x_{13}	0.0170	0.0000	0.0000	0.0000
الحالة الزوجية	x_{14}	0.0035	0.0000	0.0000	0.0000

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

الوظيفة	x_{15}	-0.0777	0.0000	0.0000	-0.01105
التحصيل الدراسي	x_{16}	0.0604	0.0203	0.27963	0.03522
حالة المريض عند الخروج	x_{17}	0.4714	0.4018	0.780739	0.42111

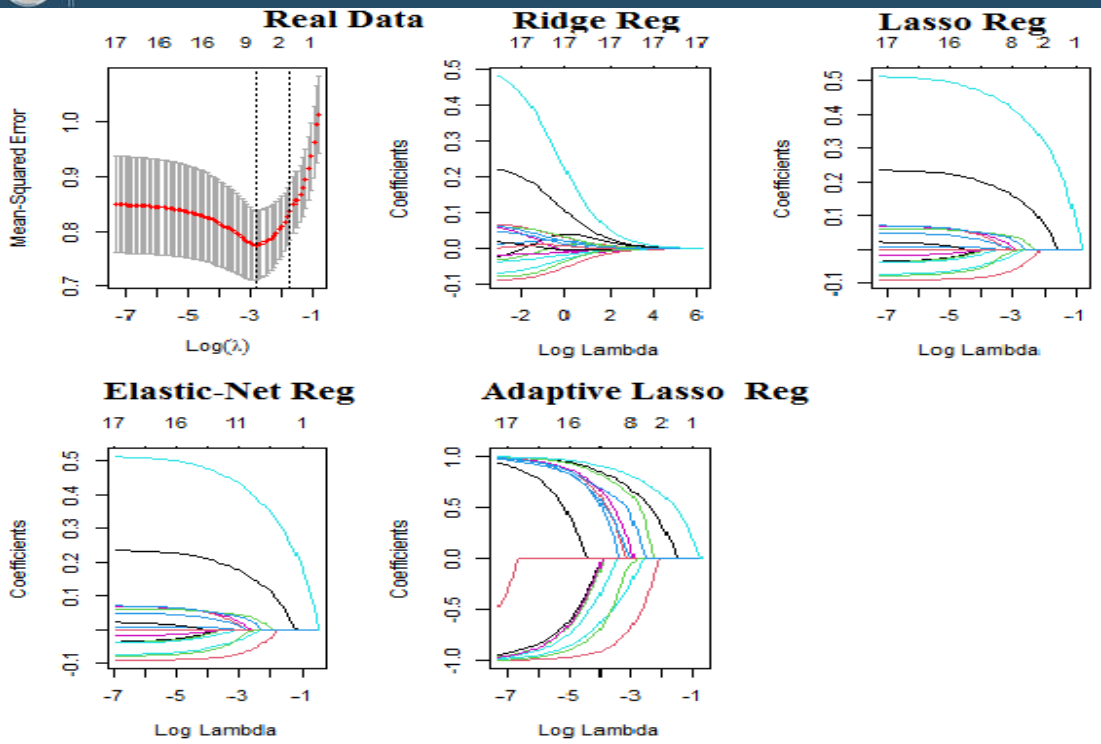
يتبين من الجدول اعلاه ما يلي:

ان طريقة (lasso, adaptive lasso, elastic net) قد عملت كأسلوب اختيار المتغيرات وهذا يتطابق مع طبيعتها النظرية حيث انها تعمل على تصفير معالم المتغيرات اتي ليس لها تأثير على متغير الاستجابة. و ان المتغيرات (العمر, الجنس, الوزن, التدخين, داء السكري, عدوا كوفيد, مدة الرقود, سكن المريض, محل الاقامة, الحالة الزوجية, الوظيفة) ليس لها اي تأثير على متغير الاستجابة حيث كانت معالم هذه المتغيرات تساوي صفر حسب نتائج طرائق التقدير(lasso, adaptive lasso, elastic net). و ان المتغيرات (الضغط, دي دايمر, الكوليسترول, الرجفان الأذيني , التحصيل الدراسي, حالة المريض عند الخروج) لها تأثير على متغير الاستجابة حسب نتائج طرائق التقدير(lasso, adaptive lasso, elastic net). و اما طريقة ridge فقد اظهرت ان المتغيرات (العمر, الوزن, الضغط, الكوليسترول, مدة الرقود, سكن المريض, الوظيفة) لها تأثير سلبي على حجم الجلطة بينما بقيت المتغيرات اظهرت تأثير ايجابي تأثير على متغير الاستجابة.

في ادناه جدول يوضح مسار حل معالم المتغيرات التفسيرية للطرائق (ridge, lasso, adaptive lasso, elastic net) ورسم سلوك حل معلمة الانكماش

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq



شكل رقم (١): سلوك حل المعالم للبيانات الحقيقية للطرائق (ridge, lasso, adaptive lasso, elastic net) وسلوك حل معلمة الانكماش.

من الشكل (١) أعلاه نلاحظ ان الرسم في الزاوية اليسرى العليا يتعلق بالبيانات التي تم محاكاتها والتي ترسم MSE في محور الصادات مع القيم اللوغاريتمية لمعلمة الانكماش λ. حيث تبين من هذا الرسم ان قيم MSE بدأت متناقصة ومستقرة عند زيادة قيم المعلمة ولكنها أخذت بالازدياد مع ازدياد قيم λ . أي يمكن القول ان هذا الرسم بين قيم معلمة الانكماش λ التي انتجتها طريقة K-fold Cross validation

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

التي تقابل اقل متوسط خطأ لتحقق المتقاطع Cross validation من نفس الشكل ومن نفس الرسم التالي وتحت عنوان Ridge Reg الذي بين سلوك حل طريقة Ridge ويسمى هذا الرسم Ridge trace . أي ان هذا الرسم يقوم برسم المعالم المقدرة بمعلومية قيم معلمة الانكماش λ . حيث ان كل خط يمثل قيم المعالم المقدرة الحاصلة بكل متغير في النموذج . حيث تلاحظ انه عند زيادة قيم λ نتيجة القيم المقدرة للمعالم لتكون قريبة للصفر. الرسم تحت عنوان lasso Reg بين أيضا سلوك حل طريقة lasso والمسمى lasso trace . حيث يتبين ان هناك حلول اقتربت من الصفر عند قيم مبكرة من المعلمة λ وتطابقها مع القيمة صفر كلما زادت قيمة λ وهذا ينطبق مع الجانب النظري لطريقة لاسوالتي تدل على تصغير الحلول من حيث ان هناك مجموعة من حلول المعالم للمتغيرات لا تقرب للصفر ألا بزيادة قيم λ. كذلك في الرسم تحت عنوان elastic net Reg تجد ان الحلول تنطبق مع الحلول المستحصل عليها من طريقة lasso . وهذا يعكس إمكانية طريقة elastic net للحصول على الحلول الصفرية. ان الرسم تحت عنوان adaptive lasso Reg بين سلوك حل هذه الطريقة في تقدير قيم معالم المتغيرات . حيث نلاحظ أيضا ان هذه الطريقة تعمل كطريقة لاختيار المتغيرات من خلال ملاحظة سلوك بعض خطوط الرسم واقترابها للحل الصفري .

الجدول التالي يوضح قيم المعايير (AICc, AIC, BIC) الخاصة بتقييم اداء طرائق التقدير. جدول رقم (٢): قيم المعايير AIC , BIC , AICC لطرائق التقدير

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

KUFA STUDIES CENTER JOURNAL
UNIVERSITY OF KUFA
مجلة مركز دراسات الكوفة - جامعة الكوفة
KSCJ OPEN ACCESS
ISSN (P) : 1993-7016
ISSN (E) : 2708-728X
www.journal.uokufa.edu.iq

Sample size	Methods	AIC	BIC	AICc
248	Ridge Reg	700.5219	787.6549	706.2811
	Lasso Reg	678.4235	681.9369	678.4397
	Adaptive Lasso Reg	683.1849	686.6983	683.2012
	Elastic-Net Reg	670.5710	674.0844	670.5873

من الجدول (٢) يتبين افضلية طريقة elastic net بسبب ان قيم المعايير الثلاثة كانت اقل ما يمكن لهذه الطريقة وتليها adaptive lasso وتليها lasso ومن ثم ridge. اي ان طريقة elastic net وفرت افضل نموذج من الناحية التفسيرية.

الاستنتاجات (Conclusions)

أفضلية أداء طريقة الشبكة المرنة elastic net عند مختلف تجارب المحاكاة تحت مختلف حجوم العينات وهذا واضح من خلال قسم المعيار MMAD . وكذلك أفضلية هذه الطريقة في تحليل البيانات الحقيقة . و ان طرائق جزاء Lasso وما يتعلق بها من طرائق جزاء تغير طرائق رائعة من ناحية أنها تعمل وفي وقتا واحد كأسلوب اختبار المتغيرات ولتقليص تقدير المعالم .و ان طريقة الجزاء elastic net أظهرت مرونة بالتعامل من مشكلة التعدد الخطي وعند زيادة الارتباطات بين المتغيرات التوضيحية في كل من الجانب التجريبي والجانب العملي للبيانات الحقيقية .و تبين لنا من تحليل البيانات الحقيقية ان حجم الجلطة الدماغية لا يتأثر بالمتغيرات الديموغرافية للمريض (العمر ، الجنس ، سكن المريض ، محل الإقامة ، الحالة الزوجية ، الوظيفة) و ان المتغيرات الفسيولوجية (الضغط ، دي دايمر ، كولسترول ، الرجفان الاذيني) لها تأثير مباشر على حجم الجلطة الدماغية . و التعمق بالنظرية الإحصائية فيما يخص طرائق جزاء لأسو وتطوير طرائق ذات علاقة بها كأساليب الاختبار المتغيرات . و أجراء دراسة لمقارنة الطرائق المذكورة في هذه الرسالة وفق أسلوب بيز وتحت افتراض وجود توزيع لاحق معمم Generalized p

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

.و اعتماد هذه الرسالة لدى وزارة الصحة ليتم الاستفادة منها للوقوف على اهم المتغيرات والعوامل والمؤثرات التي تؤثر على ظاهرة طبية معينة بهدف منع او تطور هذه الظاهرة ومعالجتها قبل تطورها . و نقترح كذلك تطبيق هذه الطرائق على تجربة طبيعية يكون فيها عدد المتغيرات التفسيرية اكبر من المشاهدات المدروسة سوآءا وفق الأسلوب التقليدي أو وفق أسلوب غيره . و لتجنب من خطر الاصابة بالجلطة الدماغية اتباع نظام غذائي صحي ، السيطرة على ضغط الدم وداء السكر والكولسترول ودي ديمر ، ممارسة الرياضة ، التخلص من التدخين والكحول ، التخلص من الوزن الزائد .

المراجع:

[Ann Kirkland , Lisa, (2014),"LASSO Simultaneous shrinkage and selection via L_1-norm" , A thesis in the Mathematical Statistics,University of Preroria, Faculty of Natural and Agricultural Sciences.

[٢] Chatterjee, S., & Hadi, A. S. (2006). Regression analysis by example. John Wiley & Sons.

[٣] Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technimetrics, 12(1):55-67.

[٤] Li, C., Pak, D., & Todem, D. (2020). Adaptive lasso for the Cox regression with interval censored and possibly left truncated data. Statistical Methods in Medical Research, 29(4), 1243-1255.

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.

[٥] Liu, Y. (2022). Adaptive lasso variable selection method for semiparametric spatial autoregressive panel data model with random effects. Communications in Statistics-Theory and Methods, 1-19.

[٦] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused LASSO. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(1):91-108.

[٧] Tibshirani. R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological). 267-288.

[٨] Zhao, P., Rocha, G. and Yu. B. (2009). The composite absolute penalties famil for grouped and hierarchical variable Selection. The Annals of Statistics, 37(6A):3468-3497.

[٩] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2):301-320.

[١٠] Zou, H. and Zhang, H.H. (2009).On the adaptive elastic-net with a diverging number of parameters. The Annals of Statistics, 37(4):1733-1751.

[١١] Zou, H.(2006). The adaptive LASSO and its oracle properties. Journal of the American Statistical Associa-fion, 101(476):1418-1429.

This work is licensed under a https://creativecommons.org/licenses/by/4.0 International License.
مجلة مركز دراسات الكوفة - جامعة الكوفة مرخصة بموجب ترخيص المشاع الإبداعي 4.0الدولي.