NTU Journal for Administrative and Human Sciences / Special Issue
DOI: https://doi.org/10.56286/ntujahs

IRAQI Academic Scientific Journals

JAHS NTU

P-ISSN: 2789-1240 E-ISSN:2789-1259
NTU Journal for Administrative and Human Sciences
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JMS/index

Employing Artificial Intelligence for Bandwidth Selection in the Multivariate Nadaraya-Watson estimator

1st, Zeina Ameer Hadied [1], 2nd, Oday Esam Al-Saqal [2] , 3rd, Dr.Zakariya Yahya Algamal [3]

1st . University of Mosul, Presidency of the University of Mousl
2nd. University of Mosul, Department of Sharia
3rd. University of Mosul, Department of Statistics & Informatics

Article Informations

Received :15, 07, 2024
Accepted : 14, 09, 2024
Published online : 01, 11, 2024

Corresponding author :
Name Zeina Ameer Hadied
University of Mosul/ Presidency of the University of Mousl
Email:
zena_saeed@uomosul.edu.iq

Key Words:
kernel estimator;
smoothing matrix,
Multivariate Nadaraya-Watson estimator,
modified crow search algorithm.

A B S T R A C T

Regression analysis is becoming more and more of a topic in most research, particularly those in the fields of economics and medicine. One of the most significant and well-known regression models utilized in recent years, which have seen significant expansion, particularly in the economic and environmental aspects, is the multiple nonparametric regression model. One of the key estimators in the multiple nonparametric regression model is the multivariate Nadaraya-Watson estimator. This estimator, in turn, depends on a matrix of parameters known as smoothing parameters to estimate the multiple nonparametric regression model. Estimating these parameters is crucial to obtaining a decent fit for the estimated curve in the multiple nonparametric regression model. In this study, it was suggested to use a modified crow search algorithm as a method inspired by nature to estimate the smoothing parameter matrix (Bandwidth matrix) in the Ndaria-Watson multiple estimator. A variety of distinct nonparametric regression models were employed to generate data using the Monte Carlo simulation approach. The mean square error was used as a benchmark for comparison, and the simulation results demonstrated the superiority of the suggested method over alternative estimating techniques.

توظيف الذكاء الاصطناعي لأختيار عرض الحزمة في مقدر نداريا- واتسون المتعدد

م.م زينة أمير بشير | م.م.عدي عصام الصقال | أ. د. زكريا يحيى الجمال

رئاسة جامعة الموصل | كلية العلوم الاسلامية | كلية علوم الحاسوب والرياضيات

المستخلص

يحظى موضوع تحليل الانحدار باهتمام متزايد وواضح في معظم الدراسات، وخاصة الاقتصادية منها والطبية. يعد نموذج الانحدار اللامعلمي بشكل عام ونموذج الانحدار اللامعلمي المتعدد بشكل خاص من أهم وأبرز نماذج الانحدار المستخدمة في السنوات الأخيرة والتي شهدت توسعا كبيرا خاصة في الجوانب الاقتصادية والبيئية. يعد مقدر Nadaraya-Watson متعدد المتغيرات أحد أهم المقدرات المستخدمة في نموذج الانحدار غير المعلمي المتعدد. في تقدير نموذج الانحدار اللامعلمي المتعدد، يعتمد هذا المقدر بدوره على مصفوفة من المعلمات تسمى معاملات التجانس، والتي يعتبر تقديرها ذا أهمية كبيرة في تحقيق التوافق الجيد للمنحنى المقدر في نموذج الانحدار اللامعلمي المتعدد. في هذا البحث تم اقتراح استخدام خوارزمية مستوحاة من الطبيعة في عملية تقدير مصفوفة عرض النطاق الترددي في مقدر نداريا-واتسون المتعدد. تم استخدام طريقة محاكاة مونت كارلو أيضًا لتوليد البيانات باتباع عدد من نماذج الانحدار اللامعلمية المتعددة. وأظهرت نتائج المحاكاة تفوق الطريقة المقترحة مقارنة بطرق التقدير الأخرى، وذلك باستخدام متوسط مربع الخطأ كمعيار للمقارنة.

الكلمات المفتاحية: مقدرات النواة، مصفوفة عرض الحزمة، مقدر نداريا - واتسون المتعدد، الخوارزميات المستوحاة من الطبيعة.

1- المقدمة

إنّ تحليل الانحدار(Regression Analysis) يعد أحد الأساليب الإحصائية المهمة لدراسة العديد من الظواهر الطبيعية والاجتماعية والاقتصادية والطبية وغيرها، إذ يستخدم في تمثيل العلاقة بين المتغيرات العشوائية المختلفة بالنسبة إلى عينة معينة أو بالنسبة إلى المجتمع على هيئة معادلة إحصائية لتحقيق الكثير من الأهداف المهمة التي يتوصل إليها من خلال تلك العلاقة. تعد أساليب الانحدار مفيدة في عملية بناء النماذج الإحصائية، إذ تصنف نماذج الانحدار إلى صنفين أساسيين بحسب طبيعة البيانات: نماذج الانحدار المعلمي (Parametric Regression Models) ونماذج

الانحدار اللامعلمي (Nonparametric Regression Models) (Rencher, 2002). ان نماذج الانحدار اللامعلمي (Nonparametric Regression Models) تقوم على إيجاد العلاقة بين متغير الاستجابة والمتغيرات التوضيحية من خلال منحنى يصف تلك العلاقة، لذا فإن الباحث يكون مهتماً بإعطاء وصف عام للعلاقة وليس دراسة التفاصيل الدقيقة للعلاقة في الانحدار اللامعلمي. وإن دالة العلاقة تكون غير معروفة وهذه النماذج تكون أكثر مرونة ولاتعتمد على فروض سابقة كما في الانحدار المعلمي، بل تعتمد بشكل أساسي ومباشر على البيانات (Data) حيث إن نوع البيانات يفسر الشكل الفعلي لمنحنى الإنحدار (محمد، 2011).

إن تحديد مصفوقة عرض الحزمة (Bandwidth) أو ما تسمى بمصفوفة معلمات التمهيد (H) في مقدر مقدر نداريا - واتسون المتعدد لذات أهمية كبيرة في تحديد وتقريب شكل دالة الانحدار اللامعلمي الى الدالة الأصلية، من خلال إيجاد الطريقة المثلى للموازنة بين التباين والتحيز. فعندما تكون القيمة لمعلمة التمهيد صغيرة فإن التحيز يكون صغيراً، والتباين يكون كبيراً، وهذا بدوره يؤدي إلى خشونة شكل الدالة وبعكسه نحصل على تعظيم التحيز، وتصغير التباين ويكون شكل الدالة أكثر تمهيداً (سلاسة) اذ انّ عملية الاختيار الجيد لقيم مصفوفة معلمة التمهيدية (H) يتم من خلال المفاضلة بين التحيز والتباين للحصول على أقل متوسط مربعات خطأ (MSE).

2- هدف البحث وأهميتهُ

ان هذا البحث يهدف الى توظيف الذكاء الاصطناعي من خلال توظيف خوارزميات التقنيات الذكائية وهي خوارزمية الغراب المعدلة (modified crow algorithm) لتقدير قيم مصفوفة معلمة التمهيد (H) بحيث تكون أكثر كفاءة مقارنة مع الطرائق الأخرى، تعمل على تحسين النتائج في عملية تقدير قيم المصفوفة (H) من خلال تجارب المحاكاة وبأستعمال نماذج مختلفة. وأما اهمية هذا البحث فتكمن في كونه يسلط الضوء على أهمية تطبيق بعض الخوارزميات الذكائية في تقدير مصفوفة عرض الحزمة وتصنيفها كأحدى الطرائق البديلة للطرائق الأحصائية التقليدية التي تخص الموضوع.

3- المقدرات اللامعلمية Nonparametric estimator

إن المرونة العالية التي تتمتع بها المقدرات اللامعلمية نظراً لكونها لا تتطلب توفر فروض بشأن توزيع المجتمع قياساً بالمقدرات المعلمية والتي تتطلب مجموعة من الفروض، ونظراً للتطور الهائل

في أجهزة الحواسيب أدى إلى ميل الباحثين في العقود الأخيرة للأهتمام بموضوع الأنحدار اللامعلمي، وطرائق التمهيد االخاصة بهِ، ومن أنواعه: نموذج الأنحدار اللامعلمي البسيط (simple Nonparametric Regression Model) والذي يقوم على إيجاد العلاقة بين متغير الاستجابة ومتغير توضيحي واحد فقط، وصيغته كما يأتي: (Koláček & Horová, 2017)

$$\mathbf{y}_i = m(\mathbf{x}_i) + \boldsymbol{\varepsilon}_i \qquad i = 1,2,\ldots,n \tag{1}$$

أما إذا كان النموذج يقوم على إيجاد العلاقة بين متغير الإستجابة، وعدد من المتغيرات التوضيحية، فعندئذٍ يسمى بنموذج الانحدار اللامعلمي المتعدد (Multivariate Nonparametric Regression Model) إذ أن هذه العلاقة تكون غير معروفة ويتم تقديرها باستعمال طرائق عدة منها: طريقة التقدير اللُبي (Kernel estimation)، وطريقة الشرائح التمهيدية (Smoothing Splines)، وطريقة المويجة (Wavelet estimator)، وصيغة معادلته تكونبالشكل الآتي (محمد و عبدالحسن، 2018;عيسى ومناف،2012):

$$\mathbf{y}_i = \mathbf{m}(\mathbf{x}_{1i}, \mathbf{x}_{2i}, \ldots, \mathbf{x}_{di}) + \boldsymbol{\varepsilon}_{\mathbf{i}} \qquad i = 1,2,\ldots,n \tag{2}$$

حيث أن : تمثل مصفوفة متجهات المشاهدات للمتغيرات التوضيحية.

وأختصاراً تكتب:

(3)

: تمثل دالة الأنحدار غير المعروفة والمطلوب تقديرها بالطرائق اللامعلمية.

هناك العديد من الطرائق اللامعلمية لتقدير هذه الدالة غير المعروفة والموضحة في المعادلة (1) و(2)، إذ إن الهدف من التمهيد هو لتقريب دالة الانحدار اللامعلمي التقريبية الى دالة الانحدار اللامعلمية الحقيقية، والعمل على تعديل المشاهدات، إذ أنّ هذه الطرائق بنيت على أساس نموذج مقدر؛ ليعطي نموذجاً مقارباً للواقع وللتنبؤ بالمستقبل، وإن معظم الطرائق اللامعلمية تفترض أن الخطأ يتوزع بمتوسط مساوٍ للصفر وتباين محدد وإن دالته هي دالة مستمرة (Continuous Function) وممهدة (Smoothed)، ومن المقدرات اللامعلمية الإحصائية التي لاتتطلب توفر فروض بشأن توزيع المجتمع، والذي يعد أداة فعالة تعتمد بشكل أساسي على البيانات (Data) هو مقدر مقدر ندارايا-واتسون (Nadaraya-Watson estimator). (Hardle, 1994)

4- مقدر نداريا-واتسون المتعدد Multivariate Nadaraya- Watson Estimator (MNWE)

أن مقدر Nadaraya-Watson يُعد من أكثر المقدرات الشائعة الاستخدام في تقدير دالة الأنحدار اللامعلمي، تم اقتراحه هذا المقدر عام (1964) من قبل الباحثين Nadaraya وWatson، كما أنه يعد من أبسط أنواع الممهدات، إذا غالباً ما يستعمل على في العديد من المجالات البحثية الإحصائية بالأعتماد على طريقة متسلسلة الأوزان كما في الصيغة (4)، إن عملية تقدير دالة الانحدار اللامعلمي $m(\mathbf{X})$ غير المعروفة تتم باستخدام المتوسط الموزون،(Aydin, 2007)

وتعرف طريقة المتوسط الموزون بأنها مشابهة لطريقة المربعات الصغرى الموزونة وكما هو مبين في المعادلة الآتية :

$$\hat{m}(\mathbf{x_i}) = n^{-1}\sum_{i=1}^{n}\mathbf{w}_i(\mathbf{x_i})\mathbf{y_i}\ ; \qquad i = 1,2,\ldots,n \qquad \textbf{(4)}$$

إذ أن $\mathbf{w}_i(\mathbf{x}_i)$: تمثل سلسلة من الأوزان الطبيعية الموجبة التي تعتمد على كل قيم المتجه وإن مجموع هذه الأوزان يساوي الواحد (Boente, et al.,1997)

$$\sum_{i=1}^{n}\mathbf{w}_i(\mathbf{x}_i) = 1 \quad ; \quad w_i \geq 0 \qquad \textbf{(5)}$$

وإن دالة الوزن لها الخصائص الآتية: (Hardle, 1994)

(1

2) $\mathbf{x} > 0$ إذا كانت دالة الوزن $\mathbf{w}(\mathbf{x})$ متزايدة

3) $\mathbf{w}(\mathbf{x}) = 0 \quad for\ |\mathbf{x}| \geq 1$

إذ إن هذه الأوزان تمثل دالة المسافة في فضاء X، والصيغة العامة للأوزان تكتب بالشكل الآتي :(Hardle, 1990& Aydin, 2007)

$$\mathbf{w}_i(\mathbf{x_i}) = \frac{k_h\ (\mathbf{x_i} - x_0)}{\hat{f}_h\ (x)} \qquad i = 1,2,\ldots,n \qquad \textbf{(6)}$$

وان:

$$\hat{f}_h(\mathbf{x}) = n^{-1}\sum_{i=1}^{n} k_h\ (\mathbf{x_i} - x_0) \qquad \textbf{(7)}$$

حيث:

$$k_h\ (u) = h^{-1}k\,(u\ /\ h\ ) \qquad \textbf{(8)}$$

إذ أن:

h: المعلمة التمهيدية (Smoothing Parameter)، أو عرض الحزمة Bandwidth، وتكون قيمتها أكبر من الصفر.

حيث إن $k(u/h_n)$ تمثل إحدى الدوال اللُبية، كما إن متسلسلة الأوزان للدالة اللُبية يشار إليها بالمختصر $\{\mathbf{w}_i\,(\mathbf{x_i})\}_{i=1}^{n}$، وتتمثل بالأوزان المؤشرة للمشاهدات i بالنسبة لـ التي تعتمد على المسافة بين النقطة x_i والنقطة x_0، إذ عادة ما تكون هذه الأوزان كبيرة إذا كانت المسافات قليلة، وتقل في حالة كون المسافات كبيرة، وأبسط طريقة لتمثيل دالة الأوزان هذه هي بوصف شكل دالة الأوزان $\mathbf{w}_i(\mathbf{x})$ بدالة الكثافة وبمعلمة ثابتة، والتي بدورها تقوم بتعديل حجم الأوزان بالقرب من النقطة x_0، وبالتالي سيكون المقدر بالصيغة الآتية :(Hardle, 1990)

$$\hat{m}\ (\mathbf{x}) = \frac{\sum_{i=1}^{n} k_h(\mathbf{x}_i - x_0)\mathbf{y}_i}{\sum_{i=1}^{n} k_h(\mathbf{x}_i - x_0)} = \frac{k\,(u)}{\sum k\,(u)} \qquad \textbf{(9)}$$

أي يصبح شكل المقدر بالصيغة الآتية:

$$\hat{m}\ (\mathbf{x}) = \frac{\sum_{i=1}^{n} k\,(\frac{\mathbf{x}_i - x_0}{h})\mathbf{y}_i}{\sum_{i=1}^{n} k\,(\frac{\mathbf{x}_i - x_0}{h})}, \qquad h > 0 \qquad \textbf{(10)}$$

وبتعميم ذلك في حالة وجود أكثر من متغير توضيحي أي d من المتغيرات التوضيحية تكون الصيغة لمقدر Nadaraya-Watson كالآتي (Soméa & Kokonendjia, 2015):

$$\hat{m}(\mathbf{x})_{MNW} = \frac{\sum_{i=1}^{n} \mathbf{K}\left(\mathbf{H}^{-1}\left(\mathbf{X_i} - \mathbf{x_0}\right)\right)\mathbf{y_i}}{\sum_{i=1}^{n} \mathbf{K}\left(\mathbf{H}^{-1}\left(\mathbf{X_i} - \mathbf{x_0}\right)\right)} \qquad \textbf{(11)}$$

حيث إن K(.) : تمثل الدالة اللُّبية المتعددة.

H: مصفوفة عرض الحزمة ذات بعد ($d \times d$) وتكون قطرية ومتماثلة و موجبة التعريف.

أي إن:

$$,\mathbf{h} = \left[h_1, h_2, \cdots, h_d\right]' \qquad \textbf{(12)}$$

(13)

5- عرض الحزمة Bandwidth

إن عملية اختيار عرض الحزمة (Bandwidth) تعد الخطوة الأكثر أهمية في تقريب دالة الانحدار اللامعلمي إلى الدالة الأصلية، ومن اجل الحصول على التقريب الجيد والملائم لا بد من إيجاد الطريقة المُثلى لأجل الموازنة بين كل من التباين و التحيز بحيث يكون مقدار الخطأ أقل مايمكن والذي عادة يقاس بمعيار متوسط مربعات الخطأ Mean Squared Error(MSE) أو متوسط مربعات الخطأ التكاملي Mean Integrated Squared Error(MISE) تتأتى هذه الموازنة من خلال استخدام أفضل قيمة لعرض الحزمة، إذ إن اختيار هذه القيمة يجب أن يكون بعناية وحذر، وذلك لكون القيمة الصغيرة جداً تؤثر على تمهيد المنحنى، وتكوّن منحنى تمهيد منخفض (Under Smoothing) ، وتكوّن منحنى تمهيد مرتفع(Over Smoothing) في حالة كانت القيمة كبيرة جداً .(Schimek, 2013& Chn ,1995)

ويرمز لمعلمة عرض الحزمة بالرمز(h) إذا كانت تستخدم لأحادي المتغير، حيث تتضمن اختيار معلمة مفردة، أما في حالة متعدد المتغيرات فتتضمن اختيار مصفوفة عرض الحزمة ويرمز لها بالرمز(H)، وتوجد عدة أساليب لاختيار القيمة المُثلى لعرض الحزمة (Bandwidth) والتي تحاول تقليل مجموع مربعات الخطأ للنموذج وهي: طريقة العبور الشرعي(Cross Validation(CV)،

وطريقة العبور الشرعي العام Generalized Cross Validation(GCV) وغيرها من الطرائق الأخرى(Mustafa & Algamal; 2022).

6- طريقة العبور الشرعي Cross Validation(CV)

هذه الطريقة تُعد من الطرائق شائعة الأستخدام؛ لإيجاد أنسب قيمة للمعلمة التمهيدية h، إذ تلعب هذه القيمة دور مهماً في تباين وتحيزالمقدر. إن فكرة هذه الطريقة تقوم على أساس تقسيم البيانات الى L من المجاميع الجزئية $(g_1, g_2,\ldots,\mathrm{g}_l)$ بحيث أن كل مجموعة تحتوي على عدد متساوي من المشاهدات $n_j = (n_1, n_2,\ldots, n_l)$، إذ يتم استبعاد مجموعة واحدة في كل مرة وتكون المجموعة المستبعدة هي g_j بحيث أن $j = 1,2,\ldots,l$ ، و كحالة خاصة عندما L=1 تسمى Leave One Out Method، أما في حالة وجود أكثر من متغير توضيحي أي d من المتغيرات التوضيحية نستعمل الصيغة كالآتي: (Hardle, 1990 & (Koláček & Horová, 2017):

$$CV(\mathbf{H}) = \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{y_i} - \hat{m}_{-1}(\mathbf{X_i};\mathbf{k})\right]^2 \qquad \textbf{(14)}$$

$\hat{m}_{-j}(\mathbf{x}_i)$: تمثل مقدرات Nadaraya-Watson في حالة استبعاد مشاهدة.

ولإيجاد مصفوفة قيم المعلمات التمهيدية المُثلى H نستعمل الصيغة الآتية:

$$\hat{\mathbf{H}} = \arg\min_{H\in H} \mathbf{H} \qquad \textbf{(15)}$$

7- طريقة العبورالشرعي العام Generalized Cross Validation (GCV)

هذه الطريقة GCV مستمدة من طريقة العبور الشرعي CV، حيث يتم الحصول عليها من صيغة CV الموضحة في الصيغة (14) وذلك عن طريق استبدال عناصر القطر الرئيسي $\hat{m}(x_i)$ لمصفوفة التمهيد S_h بمعدلها أي إن (Kauermann, & Opsomer, 2004,; Mustafa& Algamal;2022)

$$GCV_{\mathbf{H}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\hat{m}\left(\mathbf{x_i}\right) - \mathbf{y_i}}{1 - tr\left(S_{\mathbf{H}}\right)/n}\right\}^2 \qquad \textbf{(16)}$$

ولايجاد المصفوفة **H** الخاصة بقيم المعلمات التمهيدية المُثلى نستعمل الآتي:

0 (17)

8- خوارزمية الغراب المعدلة: modified crow search Algorithm

تعد خوارزمية البحث عن الغراب واحدة من أحدث الخوارزميات التطورية المستوحاة من السلوك الاجتماعي للغراب. تم تقديم هذه الخوارزمية في عام 2016 بواسطة Askarzadeh . في CSA ، يتم تحفيز الفكرة من عملية تخزين الطعام الزائد في أماكن الاختباء ثم استعادتها في الوقت اللازم. من المعروف أن الغراب طائر ذكي للغاية يراقب الآخرين وهم يخفون طعامهم ويسرقونه بمجرد مغادرتهم. بعد ارتكاب السرقة ، تختبئ لتتجنب الوقوع ضحية لها في المستقبل.

ليكن لدينا قطيع n_c من الغربان وكل غراب i له موقع عند التكرار t هو x_i^t. يتم حفظ مكان اختباء الطعام الذي يتبعه الغراب. يتحرك في مستوى البحث ويحاول العثور على أفضل مصدر للطعام والذي يعرف

بـــ M_i^t. ان نهج البحث في CSA له سيناريوهان محتملان ؛ الأول هو أن مالك الغراب مصدر الغذاء M_j^t لا يعرف أن الغراب السارق j يتبعه لذلك يصل الغراب اللص إلى مكان اختباء مالك الغراب. تتم عملية تحديث موضع لص الغراب بواسطة

(18)

حيث ان fl تمثل طول الطيران وان τ هي رقم عشوائي ضمن الفترة 0 و 1.

أما السيناريو الثاني هو أن مالك الغراب يعرف أن غراب اللص يتبعه ، لذلك فإن الغراب المالك سوف يخدع الغراب بالذهاب إلى أي موقع آخر في مساحة البحث. يتم تحديث موضع الغراب بواسطة موضع عشوائي. في CSA ، يتم تحديد السيناريو من خلال التعبير التالي:

(19)

حيث ان θ هي رقم عشوائي ضمن الفترة 0 و 1. اما AP هي احتمالية الادراك.

لإجراء اختيار المتغير ، تم اقتراح خوارزمية ثنائية للبحث عن الغراب (Sayed, Hassanien, & Azar, 2017). على عكس CSA القياسي ، حيث يتم تحديث الحلول في مساحة البحث نحو

المواضع ذات القيمة المستمرة ، في BCSA ، تم تصميم مساحة البحث على شكل شبكة منطقية ذات أبعاد n ويتم تحديث الحلول عبر زوايا المكعب الفائق. بالإضافة إلى ذلك ، نظرًا لأن المشكلة تكمن في اختيار أو عدم تحديد متغير معين ، يتم استخدام متجه ثنائي للحل ، حيث يتوافق 1 مع ما إذا كان سيتم تحديد متغير لتكوين مجموعة البيانات الجديدة ، و 0 بخلاف ذلك. في أي خوارزمية ثنائية ، حيث يستخدم المرء متجه الخطوة لحساب احتمالية تغيير المواضع ، تؤثر وظائف التحويل بشكل كبير على التوازن بين الاستكشاف والاستغلال.

(Islam & Mei,2017),(Mafarja, et al, 2018)

في BCSA ، تُستخدم وظيفة النقل لتعيين مساحة بحث مستمرة إلى مساحة ثنائية ، وتم تصميم عملية التحديث لتبديل مواقع النجوم بين 0 و 1 في مساحات البحث الثنائية. من أجل بناء هذا المتجه الثنائي ، وظيفة النقل في المعادلة. (20) يمكن استخدامها ، حيث يكون الحل الجديد مقيدًا بالقيم الثنائية فقط

$$x_i^t = \begin{cases} 1 & \text{if} \quad T(x) > \alpha \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{(20)}$$

حيث ان $\alpha \in [0,1]$ هي عبارة عن رقم عشوائي وان $T(x)$ هي دالة تحويل. ان دالة التحويل تعرف بالشكل الاتي:

(21)

في هذا البحث تم اقتراح استخدام دالة تحويل متغيرة خلال الزمن. اي ان دالة التحويل هذه سوف تتغير خلال تكرار الحل. تم هذا الاقتراح من خلال اضافة معلمة تحكم وهي θ، اذا تحتاج هذه المعلمة الى قيمة عليا وقيمة دنيا لها من خلال المعادلة الخاصة بها وهي:

(22)

وعليه سوف تصبح دالة التحويل المقترحة بالشكل التالي:

(23)

من أجل إتمام هدف البحث وتحقيقهُ، وبالاعتماد على هذه التقنية، فإن كل عنصر (غراب) في المجموعة سيكون لديه d من المواقع التي تمثل عدد المتغيرات التوضيحية في انموذج انحدار كاوس المعكوس. بناءً على ذلك، فإن توظيف خوارزمية الغراب تكون وفق الخطوات التالية:

الخطوة الأولى: تحديد حجم المجموعة (عدد الغربان) وهو 30 غراب، حيث إن كل غراب سيكون له متجه من عدد المتغيرات المستقلة فضلاً عن ذلك تحديد عدد التكرارات داخل خوارزمية الغراب حيث استقرت النتائج عند التكرار 500.

الخطوة الثانية: توليد القيم الأولية التي تحتاجها الخوارزمية، التي ستمثل القيم الأولية الافتراضية ، فإن توليدها سيكون من التوزيع المنتظم المستمر وفق الفترة [0,1].

الخطوة الثالثة: لغرض اختيار القيم المُثلى، تم الاعتماد على Fitness Function وفق الصيغة الآتية:

$$\text{Fitness Function} = \min\left[\frac{\sum_{i=1}^{n}(\mathbf{y_i} - \hat{m}(\mathbf{X}))^2}{n}\right] \quad \textbf{(24)}$$

الخطوة الرابعة: بالاعتماد على أقل قيمة يحصل عليها أي غراب وفق المعادلة (20) يتم تحديث مواقع باقي الغربان.

الخطوة الخامسة: نستمر بالحل لحين الوصول الى أعلى تكرار للخورازمية، الذي تم تحديده بالخطوة الأولى والذي سيمثل الحل الأمثل.

x_1	x_2	....	x_{p-1}	x_p
1	0	....	1	0

الشكل 1: الية اختيار المتغيرات حسب خوارزمية الغراب المعدلة

9- نتائج المحاكاة

لاختبار مدى جودة أداء الطريقة المقترحة تم تصميم العديد من التجارب ومحاكاتها باستعمال لغة البرمجة(R) ، إذ تم الأخذ بنظر الاعتبار ثلاثة أحجام للعينات(n=(50,100,200 ، وتم إجراء المقارنات للطرق المختلفة المستخدمة CV، GCV، مع الطريقة المقترحة لخوارزمية (CSA) وباستخدام دالة Epanechnikov كدالة لُبية، مع دراسة أربعة نماذج مختلفة هي:

النموذج الأول: تم توليد هذا الأنموذج وفق المعادلة الآتية (Lijian & Rolf, 1999) :

$$\mathbf{y_i} = \left\{ \left(\mathbf{x_1} - 0.5 \right)^2 + \mathbf{x_2^2} \right\} \sin\left(2\pi \mathbf{x_3} \right) + \boldsymbol{\varepsilon}_\mathbf{i} \qquad \textbf{(25)}$$

حيث تم توليد المتغيرات التوضيحية X_1 وX_2 وX_3 من التوزيع المنتظم ضمن الفترة [0,1]، أما $\boldsymbol{\varepsilon}_\mathbf{i}$ فيتوزع بالشكل الآتي:

النموذج الثاني: تم توليد هذا النموذج وفق المعادلة الآتية (Goutte,et.al,2000) :

(26)

حيث تم توليد المتغيرات التوضيحية X_1 وX_2 وX_3 وX_4 وX_5 من التوزيع المنتظم ضمن الفترة [0,1]، أما $\boldsymbol{\varepsilon}_\mathbf{i}$ فيتوزع بالشكل الآتي :

النموذج الثالث: تم توليد هذا النموذج وفق المعادلة الآتية (Shang,et.al,2014) :

$$\mathbf{y_i} = \sin\left(2\pi\, \mathbf{x_1} \right) + 4\left(1 - \mathbf{x_2} \right)\left(1 + \mathbf{x_2} \right) + \frac{2\mathbf{x_3}}{1 + 0.8\, \mathbf{x_3^2}} + \boldsymbol{\varepsilon}_\mathbf{i} \qquad \textbf{(27)}$$

حيث تم توليد المتغيرات التوضيحية X_1 وX_2 وX_3 من التوزيع المنتظم ضمن الفترة [0,1]، أما $\boldsymbol{\varepsilon}_\mathbf{i}$ فيتوزع بالشكل الآتي:

النموذج الرابع: تم توليد هذا النموذج وفق المعادلة الآتية(Koláček & Horová, 2017):

$$\mathbf{y_i} = (\mathbf{x_1} - 0.5)^3 + (\mathbf{x_2} - 0.5) + \boldsymbol{\varepsilon}_\mathbf{i} \qquad \mathbf{(28)}$$

كما وتم توليد المتغيرات التوضيحية X_1 وX_2 من التوزيع المنتظم ضمن الفترة [0,1]، أما $\boldsymbol{\varepsilon}_\mathbf{i}$ فيتوزع بالشكل الآتي:

$$\boldsymbol{\varepsilon}_\mathbf{i} \sim N(0, 0.02)$$

حيث تم تكرار اجراء التجربة 250 مرة، وتم الأعتماد على متوسط مربعات الخطأ (MSE) كمعيار للمقارنة بين طرائق التقدير المستخدمة وبيان الطريقة الأفضل وتلخيص نتائج الطرق المستخدمة في الجداول من (1-4).

اظهرت النتائج تفوق طريقة (CSA) على باقي الطرائق الاخرى (CV,GCV) من حيث معاير MSE فعلى سبيل المثال جدول (3) اعطت طريقة (CSA) اقل قيمة عندما n=100 ، حيث بلغت0.046 مقارنة بـــ 0.314 لطريقة CV و0.294 لطريقة GCV بالاضافة الى ذلك الطريقة المقترحة حافظت على افضليتها عند تغير حجم العينة ولجميع النماذج المستخدمة.

الجدول(1): معدل قيمة MSE وقيمة الانحراف المعياري للأنموذج الأول في مقدرNadaraya-Watson

	CV	GCV	CSA
n=50	0.041 ± 0.007	0.038 ± 0.008	0.022 ± 0.005
n=100	0.038 ± 0.005	0.033 ± 0.005	0.018 ± 0.003
n=250	0.031 ± 0.002	0.029 ± 0.003	0.013 ± 0.003

الجدول(2): معدل قيمة MSE وقيمة الانحراف المعياري للأنموذج الثاني في مقدرNadaraya-Watson

	CV	GCV	CSA
n=50	3.105 ± 0.711	2.971 ± 0.920	0.0125 ± 0.0024
n=100	3.098 ± 0.636	2.812 ± 0.735	0.0057 ± 0.00022
n=250	3.055 ± 0.311	2.057 ± 0.229	0.0038 ± 0.0005

الجدول(3): معدل قيمة MSE وقيمة الانحراف المعياري للأنموذج الثالث في مقدر مقدرNadaraya-Watson

	CV	GCV	CSA
n=50	0.327 ± 0.055	0.321 ± 0.050	0.065 ± 0.041
n=100	0.314 ± 0.031	0.294 ± 0.031	0.046 ± 0.025
n=250	0.297 ± 0.017	0.271 ± 0.022	0.027 ± 0.014

الجدول(4): معدل قيمة MSE وقيمة الانحراف المعياري للأنموذج الرابع في مقدر مقدرNadaraya-Watson

	CV	GCV	CSA
n=50	0.620 ± 0.138	0.344 ± 0.088	0.0058 ± 0.0021
n=100	0.523 ± 0.117	0.347 ± 0.048	0.0047 ± 0.0015
n=250	0.341 ± 0.044	0.201 ± 0.035	0.0028 ± 0.0011

10- الاستنتاجات:

من خلال مقارنة الطريقة المقترحة CSA وطرائق التقدير اللامعلمية المعتمدة لتقدير مصفوفة معلمات التمهيد(H) في نموذج الإنحدار اللامعلمي، أن أفضل طريقة كانت الطريقة المقترحة CSA ، لتحقيقها أقل قيمة لمعيار متوسط مربعات الخطأ(MSE) ولجميع حجوم العينات الثلاثة 100, 50) (250,، ولجميع النماذج الأربعة المعتمدة في تجارب المحاكاة. كذلك جاءت طريقة GCV أفضل طريقة لتقدير مصفوفة (H) بعد الطريقة المقترحة CSA لإعطائها نتائج جيدة في تقدير مصفوفة(H)، وكانت طريقة CV بالمرتبة الاخيرة لإعطائها أعلى قيم لـ MSE مقارنة بالطرائق .

11- المصادر:

1. عيسى، أسيل مسلم و حمود، مناف يوسف،(2012)، "مقارنة بعض المقدرات شبه المعلمية لتقدير دالة الانحدار"، مجلة العلوم الاقتصادية والإدارية، المجلد (18)، العدد (67) الصفحات [273- 288] .

2.محمد، لقاء علي وعبد الحسن، ميسم عبد النبي، (2018)، "مقارنة المقدرات اللامعلمية في تحليل الانحدار المتعدد لدالتي كاما وبيتا" ، مجلة العلوم الاقتصادية والإدارية، المجلد (24)، العدد (108)، الصفحات [497- 488]

3. محمد، محمد عبد الحسين، (2011)، "استخدام مقدر كيرنل Nadaraya-Watson في تقدير دالة الانحدار اللامعلمي"، مجلة القادسية للعلوم الإدارية والاقتصادية، المجلد (13)، العدد (1) 15.

المصادر الأجنبية:

4. Askarzadeh, A., A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm. Computers & Structures, 2016. 169: p. 1-12.

5. Aydin, D., (2007), "A comparison of the nonparametric regression models using smoothing spline and kernel regression", World Academy of Science, Engineering and Technology, (36), PP[253-257].

6. Boente, G., Fraiman, R., & Meloche, J., (1997), "Robust plug-in bandwidth estimators in nonparametric regression", Journal of Statistical Planning and Inference, 57(1), pp[109-142].
7. C.K. Chn (1995), " Bandwidth selection in Nonparametric Regression With general errors", Journal of Statistic Planning and Inference, 44 ,PP. 265-275.
8. Goutte, C., Larsen, J., & technology, v., (2000), "Adaptive metric kernel regression", Journal of VLSI signal processing systems for signal, image, 26(1-2), pp[155-167].
9. Hardle, W., (1990), "Applied Nonparametric Regression", Cambridge MA : Cambridge University press.
10. Harldle , W. , (1994)," Applied non parametric regression ".
11. Islam, M.J., X. Li, and Y. Mei, A time-varying transfer function for balancing the exploration and exploitation ability of a binary PSO. Applied Soft Computing, 2017. 59: p. 182-196.
12. Kauermann, G., & Opsomer, J., (2004), "Generalized cross-validation for bandwidth selection of backfitting estimates in generalized additive models", Journal of Computational and Graphical Statistics, 13(1),PP[66-89].
13. Koláček, J., & Horová, I., (2017), "Bandwidth matrix selectors for kernel regression", Computational Statistics, 32(3), [1027-1046].
14. Lijian, Y., & Rolf, T., (1999), "Multiverait bendwidth selection for local linear regression", Statist, 61, pp[793-815].

15. Mafarja, M., et al., Binary dragonfly optimization for feature selection using time-varying transfer functions. Knowledge-Based Systems, 2018. 161: p. 185-204.
16. Mehrabian, A. R., & Lucas, C. (2006). A novel numerical optimization algorithm inspired from weed colonization. Ecological informatics, 1(4), 355-366.
17. Mustafa, M. Y. & Algamal, Z. Y.,(2022)," Bandwidth Selection in Multivariate Nadaraya-Watson Estimator based on Meta-Heuristic Optimization Algorithms: A Simulation Study" Mathematical Statistician and Engineering Applications, 71(4), [4877-4887].
18. Sayed, G.I., A.E. Hassanien, and A.T. Azar, Feature selection via a novel chaotic crow search algorithm. Neural Computing and Applications, 2017.
19. Schimek, M. G., (2013)," Smoothing and regression: approaches, computation, and application", John Wiley & Sons.
20. Shang, H. L., Zhang, X., & Shang, M. H. L., (2014), "Package bbemkr".
21. Soméa, S. M., & Kokonendjia, C. C., (2015), "Effects of associated kernels in nonparametric multiple regressions", arXiv preprint arXiv:1502.01488.