مقارنة مقدرات عرض الحزمة (معلمة التمهيد) باستخدام الدوال اللبية في تحليل المركبات الرئيسية

أ.م.د. لقاء علي محمد		الباحث : أمير علي عبود *

كلية الادار والاقتصاد – جامعة بغداد

قسم الاحصاء

المستخلص :

عند استخدام تحليل متعدد المتغيرات مع مجاميع البيانات ذات الابعاد العالية , غالباً ما نستخدم تحليل المركبات الرئيسية (PCA) لتقليص تلك الابعاد .

حيث يعمل تحليل المركبات الرئيسية (PCA) على دراسة العلاقة بين مجموعة من المتغيرات عالية الابعاد وتحويلها الى مجموعة جديدة من المركبات والتي تلخص المتغيرات المقاسة وتكون مؤهلة لتفسير معظم التباين الكلي للبيانات الاصلية ويعتمد بصورة رئيسية على حساب مصفوفة التباين المشترك او مصفوفة الارتباط , الا ان هذا التحليل يتأثر بطبيعة البيانات ولا يمكن اجراءه الا بعد تحقيق مصفوفة التباينات للشروط الخاصة بها قبل البدء بالتحليل ومن هذه الشروط هي الخطية في البيانات وبما ان بيانات الظاهرة المدروسة في هذا البحث لا تحقق هذه الصفة كونها تتصف بخاصية اللاخطية عليه تم اللجوء الى استخدام دالة (RBF KERNEL) اللبية في تحليل المركبات الرئيسية, والتي تعتمد في حسابها على معلمة التمهيد (عرض الحزمة) وقدرت في بحثنا هذا بمجموعة من الطرق هي (مقدر المربعات الصغرى للعبور الشرعي (LSCV) , مقدر العبور الشرعي المتحيز (BCV) , مقدر العبور الشرعي الممهد (SCV) , مقدر قاعدة الملئ المباشر (PI)) , وتطبيقها على البيانات الناتجة من تجارب المحاكاة لحجوم العينات المختلفة (صغيرة , متوسطة , كبيرة) , والمقارنة بين تلك المقدرات من خلال عدد المركبات الرئيسية الفعالة التي تزيد قيم الجذور الذاتية لها عن الواحد الصحيح وما يقابلها من النسبة في تفسير التباين الكلي .

الكلمات المفتاحية : KPCA , RBF KERNEL , الدوال اللبية , تقدير عرض الحزمة , BCV LSCV , , SCV , PI .

---

*بحث مستل من رسالة ماجستير قيد المناقشة .

comparing bandwidth estimatore (smooting parameter) by Using Of Kernel function in the principal Component Analysis

Abstract

Always Principal Component Analysis (PCA) used multivariate analysis with high dimensional data sets, often using Principal Component Analysis (PCA) to reduce these dimensions,

The Principal Component Analysis (PCA) based on the study of the relationship between a group of high-dimensional variables and convert them to a new group of components,The principle component analysis based on study the relation between the high dimension variable group and transfer it to new groups of components Which summarizes the measured variables and be qualified to explain the most of the contrast of the original data , and it depends mainly on the calculation of the covariance matrix or correlation matrix but this analysis is affected by the nature of the data, it can be performed only after the covariance matrix achieve to the conditions before starting their own analysis. One of these condition is the linearity of data , Since the data of the phenomenon studied in this research do not achieve this status being characterized by non-linear feature, so it was resorting to the use of Kernel functions in the analysis of the principle components ,which is based on its calculated on the beginning identification (bandwidth) estimated in our research with variety methods ,are *Least Squares* Cross Validation (LSCV), biased crossing valid (BCV), Smoothed Cross-Validation (SCV) , Direct Plug-in Rule (DPI) . comparing these estimatore through the effective principle components numbers when their eigen values increasing more than one and corresponds from the ratio in explain the Total variance .

Keywords : KPCA , RBF KERNEL, kernel function , bandwidth estimating , BCV , LSCV , SCV , PI .

1- المقدمة Introductio

يعد تحليل متعدد المتغيرات والذي يعتبر من المواضيع المهمة والاساسية في التحليل الاحصائي فالتطور التكنولوجي والعلمي في آن واحد زاد من استخدام تحليل متعدد المتغيرات لما له من اهمية في تحليل الظواهر ولكافة الابعاد والعالية منها على وجه الخصوص .

وتأتي اهمية تحليل المركبات الرئيسية من خلال مجموعة خصائص تميزها عن غيرها , ولا يمكن التعامل مع تحليل المركبات الرئيسية الا بعد تحقيق مصفوفة البيانات للشروط الخاصة بها قبل البدء بالتحليل ومن هذه الشروط هي الخطية في البيانات والتي قد نجد مجاميع من البيانات لا تحقق هذه الصفة لذا يصبح التعامل بالتحليل العادي غير مقبول لانه لا يعطي نتائج يمكن الوثوق بها فمن احدث الوسائل في التعامل مع البيانات متعدد المتغرات والتي لا تحقق شرط الخطية هي اللجوء الى الدوال اللبية بكافة انواعها والتي تعتمد في حسابها على معلمة التمهيد (عرض الحزمة) والتي تحسب بأكثر من طريقة.

2- هدف البحث : Purpose of search

دراسة وتحليل البيانات غير الخطية (Nonlinear Data) من خلال استخدام الدوال اللبية في تحليل المركبات الرئيسية (Kernel Principal Component Analysis) والتوصل الى افضل النتائج من خلال المقارنة بين معالم مختلفة للتمهيد .

3- الجانب النظري :

3-1 الدوال اللبية في تحليل المركبات الرئيسية : [5][8][9][10][11][15]

(Kernel Function Principal Component analysis)

تعتبر الدوال اللبية من المواضيع الحديثة والمهمة لما لها من تطبيقات يصعب على الطرائق الكلاسيكية التعامل معها حيث انها تستعمل نفس الفكرة الاساسية لطريقة تحليل المركبات الرئيسية (PCA) أي انها تسعى الى تحويل مجموعة كبيرة من المتغيرات الى عدد اقل بحيث تكون المتغيرات الناتجة من عملية التحويل والتي تدعى بالمركبات (المكونات أو المحاور) الرئيسية تفسر اكبر قدر من التباين الكلي للبيانات الاصلية وذلك بأستعمال الدوال اللبية (دوال كيرنل أو دوال التحويل غير الخطية) في استخراج

مصفوفة اللب (Kernel Matrix) لتحويل البيانات من الفضاء الاصلي R^d ذات الابعاد العالية الى فضاء ذات ابعاد منخفضة والمسمى بفضاء الميزة ويرمز له بالرمز F .

ولايضاح فكرة الدوال اللبية بصورة مبسطة لا بد من التطرق الى الحالة الاعتيادية وهي حالة البيانات الخطية وكيفية التعامل معها تمهيداً لفكرة الدوال اللبية وكيفية التعامل مع البيانات غير الخطية وكالاتي .

على سبيل المثال

لو كانت لدينا مجموعة من البيانات الخطية تتضمن المشاهدات $(x_i, i = 1, 2, \ldots\ldots, N)$.

$x_i \in R^d$

اذ ان :

R^d : يمثل الفضاء الاصلي للبيانات .

وبافتراض ان انحرافات القيم عن وسطها الحسابي يساوي صفر .

$$\sum_{i=1}^{N} x_i = 0$$

فيكون التعامل مع هذه البيانات بتحليل المركبات الرئيسية الاعتيادي , ويتم استخراج الجذور المميزة (λ_i) المقابلة للموجهات المميزة (b_i) من خلال حل المعادلة الاتية .

$$C\underline{V} = \lambda\underline{V} \quad \cdots\cdots\cdots\cdots \quad (1)$$

اذ أن :

C : مصفوفة التباين والتباين المشترك المحسوبة بطريقة تحليل المركبات الرئيسية الاعتيادي وفق الصيغة الاتية .

$$C = \frac{1}{N}\sum_{i=1}^{N}(x_i)\,(x_i)' \quad \cdots\cdots\cdots\cdots \quad (2)$$

وبضرب طرفي المعادلة (2-18) بـ (x_i) تصبح .

$$x_i\ C\underline{V} = \lambda\,(x_i\,\underline{V}) \quad \cdots\cdots\cdots\cdots \quad (3)$$

الا انه عندما تكون البيانات غير خطية وذات الابعاد العالية فتصبح الالية المتبعة اعلاه غير مجدية وبالتالي يتم اللجوء الى استعمال الدوال اللبية والتي تعمل على تحويل البيانات من الفضاء الاصلي R^d ذات الابعاد العالية الى فضاء ذات ابعاد منخفضة والمسمى بفضاء الميزة ويرمز له بالرمز F وكالاتي:

$$\phi : R^d \rightarrow F$$

اذ ان :

ϕ : التحويل من الفضاء الاصلي للبيانات R^d الى الفضاء F .

وعلى سبيل المثال :

$$X = \{x_1, x_2\}$$

فأن :

$$\phi(X) = \{X_1^2, X_2^2, X_1X_2, X_2X_1\} \quad \cdots\cdots \quad (4)$$

وبافتراض ان البيانات متمركزة (انحرافات القيم عن وسطها الحسابي يساوي صفر) أي ان .

$$\sum_{i=1}^{N} \phi(x_i) = 0 \quad \cdots\cdots \quad (5)$$

فتكون الصيغة العامة لايجاد المركبات الرئيسية هي :

$$\begin{aligned} \mathbf{Y_i} &= \langle \boldsymbol{V_i}\,, \boldsymbol{\phi}(\mathbf{x_i}) \rangle \\ &= \boldsymbol{b_j}\, \langle \boldsymbol{\phi}(\mathbf{x_i})\ \ \boldsymbol{\phi}(\mathbf{x}) \rangle \\ &= \sum_{i,j=1}^{p} \boldsymbol{b_i}\, \mathbf{K}(\mathbf{x_i}, \mathbf{x}) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (6) \end{aligned}$$

اذ ان :

V_i : يمثل المتجه المميز (i) والذي يقابل القيمة المميزة (λ_i) والذان يحددان ترتيب المركبة الرئيسية ($\mathbf{Y_i}$) ويحقق نفس الشروط التي تحققها المتجهات المميزة في تحليل المركبات الرئيسية الاعتيادي , والذي يحسب وفق الصيغة :

$$\boldsymbol{V_i} = \sum_{i=1}^{N} \mathbf{b_i}\ \ \boldsymbol{\phi}(\mathbf{x_i}) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (7)$$

b_i : تمثل معاملات $\phi(x_i)$.

$$\boldsymbol{V_i}'\boldsymbol{V_j} = \begin{cases} \mathbf{0} & \mathbf{if}\ \ \mathbf{i} \neq \mathbf{j} \\ \mathbf{1} & \mathbf{if}\ \ \mathbf{i} = \mathbf{j} \end{cases} \qquad \mathbf{i}, \mathbf{j} = \mathbf{1}, \mathbf{2}, \ldots\ldots, \mathbf{p} \cdots\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots (8)$$

ويتم الجذور المميزة (λ_i) المقابلة للموجهات المميزة V_i من خلال حل المعادلة رقم (1) المذكورة انفاً باستعمال مصفوفة التباينات $\bar{\mathbf{C}}$ المحسوبة بأستعمال الدوال اللبية في تحليل المركبات الرئيسية بدلاً من المصفوفة التباينات $\mathbf{C}$ المحسوبة بطريقة تحليل المركبات الرئيسية الاعتيادي .

$$\bar{\mathbf{C}}\underline{\mathbf{V}} = \boldsymbol{\lambda}\underline{\mathbf{V}} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (9)$$

حيث ان :

$\bar{\mathbf{C}}$: مصفوفة التباين والتباين المشترك المحسوبة بطريقة التحليل اللبي وفق الصيغة :

$$\bar{\mathbf{C}} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{i=1}^{N} \boldsymbol{\phi}(\mathbf{x_i})\ \ \boldsymbol{\phi}(\mathbf{x_i})' \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (10)$$

ولايجاد قيم المعاملات (b_i) الضرورية لأيجاد المتجهات المميزة (Vi) , نعوض المعادلة رقم (7) والمعادلة رقم (10) في المعادلة رقم (9) :

$$\left[\frac{\mathbf{1}}{\mathbf{N}}\sum_{\mathbf{i=1}}^{\mathbf{N}}\boldsymbol{\phi}(\mathbf{x_j})\,\boldsymbol{\phi}(\mathbf{x_j})'\right]\left[\sum_{\mathbf{i=1}}^{\mathbf{N}}\ \boldsymbol{b}_{\mathbf{i}}\ \boldsymbol{\phi}(\mathbf{x_i})\right]=\boldsymbol{\lambda}\left[\sum_{\mathbf{i=1}}^{\mathbf{N}}\ \mathbf{b_i}\ \boldsymbol{\phi}(\mathbf{x_i})\right]\qquad\cdots\cdots(\mathbf{11})$$

وبضرب طرفي المعادلة رقم (11) بـــ ($\boldsymbol{\phi}(\mathbf{x_k})$) تصبح :

$$\boldsymbol{\phi}(\mathbf{x_k})\left[\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\mathbf{j=1}}^{\mathbf{N}}\boldsymbol{\phi}(\mathbf{x_j})\,\boldsymbol{\phi}(\mathbf{x_j})'\right]\left[\sum_{\mathbf{i=1}}^{\mathbf{N}}\ \mathbf{b_i}\ \boldsymbol{\phi}(\mathbf{x_i})\right]=\boldsymbol{\lambda}\left[\boldsymbol{\phi}(\mathbf{x_k})\sum_{\mathbf{i=1}}^{\mathbf{N}}\ \mathbf{b_i}\ \boldsymbol{\phi}(\mathbf{x_i})\right]$$

وبأعادة ترتيب المعادلة اعلاه تصبح :

$$\left[\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\mathbf{i=1}}^{\mathbf{N}}\ \mathbf{b_i}\,\langle\boldsymbol{\phi}(\mathbf{x_k})\sum_{\mathbf{j=1}}^{\mathbf{N}}\boldsymbol{\phi}(\mathbf{x_j})\ \rangle\ \langle\boldsymbol{\phi}(\mathbf{x_j}),\boldsymbol{\phi}(\mathbf{x_i})\rangle\right]=\boldsymbol{\lambda}\left[\sum_{\mathbf{i=1}}^{\mathbf{N}}\ \mathbf{b_i}\,\langle\boldsymbol{\phi}(\mathbf{x_k})\ \ \boldsymbol{\phi}(\mathbf{x_i})\rangle\right]$$

$..(\mathbf{12})$

ومن خلال المعادلة رقم (12) يتبين ان ايجاد المعاملات يعتمد على حساب قيم الـ ($\boldsymbol{\phi}$) الا انه ليس من السهل حسابها , وبالتالي تم اللجوء الى مصفوفة اللب (Kernel Matrix) (k) والمبينة لاحقاً والتي تحسب وفق الصيغة الاتية:

$$\boldsymbol{K}=\langle\boldsymbol{\phi}(\mathbf{x_j}),\boldsymbol{\phi}(\mathbf{x_j})\rangle\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\ \ \cdots\cdots\cdots\cdots\cdots\cdots(\mathbf{13})$$

نحصل على :

$$\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\mathbf{i=1}}^{\mathbf{N}}\mathbf{b_i}\sum_{\mathbf{j=1}}^{\mathbf{N}}\ \ \boldsymbol{K_{kj}K_{ji}}=\boldsymbol{\lambda}\sum_{\mathbf{i=1}}^{\mathbf{N}}\mathbf{b_i}\,\boldsymbol{K_{ki}}\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(\mathbf{14})$$

وبضرب طرفي المعادلة (14) بـــ (N) نحصل على :

$$\boldsymbol{K^2b}=\boldsymbol{\lambda NKb}\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(\mathbf{15})$$

$$\boldsymbol{Kb}=\boldsymbol{\lambda Nb}\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\ \ \cdots\cdots\cdots\cdots\cdots\cdots\cdots(\mathbf{16})$$

$$\boldsymbol{Kb}=\tilde{\boldsymbol{\lambda}}\boldsymbol{b}\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\ \cdots\ \ \cdots\cdots\cdots\cdots\cdots\cdots(\mathbf{17})$$

و ان :

$$\tilde{\lambda} = \lambda N \quad \cdots\cdots\cdots\cdots \quad (18)$$

والشرط اللازم لايجاد المعاملات الجذور المميزة (b_i) هو أن تكون المصفوفة $\left[\mathrm{K} - \tilde{\lambda}\mathrm{I}\right]$ منفردة , اي إن قيمة محددها يساوي صفر .

$$\left|\mathrm{K} - \tilde{\lambda}\mathrm{I}\right| = 0 \quad \cdots\cdots\cdots\cdots \quad (19)$$

تسمى المعادلة (19) بالمعادلة المميزة (*Characteristic Equation*) وبحل هذه المعادلة يمكن ايجاد قيمة الجذور المميزة (λ_i) المناظرة للموجهات المميزة (V_i) .

<u>2-3 مصفوفة اللب (Kernel Matrix)</u> : [5][8][9][10][11][15]

وهي مصفوفة مربعة من درجة (NxN) يرمز لها بالرمز(K) وتمتلك هذه المصفوفة خاصية التماثل (<u>Symmetric</u>) أي ان $K_{ij} = K_{ji}$ لكل زوج (I , j) , كما انها مصفوفة شبهة موجبة (Sime Positive) .

ويتم حساب مصفوفة اللب من خلال استعمال الدوال اللبية والتي تعتمد في حسابها على قيمة معلمة التمهيد (عرض الحزمة) والتي يرمز لها بالرمز (h) وتكون الصيغة العامة لهذه المصفوفة هي :

$$K = \left[K_{ij} = \langle \emptyset(X_i ,\ \emptyset(X_j)\rangle\right]$$
$$= \left[K(X_i , X_j)\ \right] \quad \cdots\cdots\cdots\cdots \quad (20)$$

اذ ان :

K : تمثل مصفوفة اللب .

$K(X_i , X_j)$: تمثل الدالة اللبية .

كما ان هذه المصفوفة تحتوي على جميع المعلومات المطلوبة لحساب ازواج المسافات بين مجموعة البيانات .

ويتضمن الجدول ادناه الصيغة الرياضية لدالة (RBF KERNEL) المستعملة في هذا البحث :

جدول رقم (3-1)

الدوال اللبية المستخدمة في البحث

$Kernel$	$K(X_i\,,X_j)$	
Radial Basis Functions RBF	$exp\left(-\frac{\left\lVert x_i - x_j\right\rVert^2}{2h^2}\right)$	**h: bandwidth**

اذ ان :

h: (Bandwidth) عرض الحزمـة او تعرف بمعلمة التمهيد (Smooth Parameter).

d : درجة دالة (Polynomia)l

3-3 توسيط البيانات في فضاء الميزة [5][8][9][10][11][15]

: (Centering Data In Feature Space)

ذكرنا سابقا بأفتراض ان تكون البيانات متمركزة (أي ان انحرافات القيم عن وسطها الحسابي يساوي صفر) وكما جاء في المعادلة رقم (22 - 2) , الا انه وبصورة عامة من الصعب جداً الحصول على بيانات تتمتع بهذه الميزة وبالتالي يجب توسيط البيانات , وكما تبين أنفا ان ايجاد المركبات الرئيسية بأستعمال الدوال اللبية يعتمد بشكل رئيسي على مصفوفة اللب فبالامكان اجراء عملية التوسيط لمصفوفة اللب والذي بدوره ينعكس على البيانات بفضاء الميزة وكما يأتي :

$$\widetilde{\emptyset}(X_i) = \emptyset(X_i) - \frac{1}{N}\sum_{k=1}^{N}\phi(x_k) \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (21)$$

وبتعويض المعادلة رقم (38-2) في المعادلة رقم (37-2) نحصل على :

$$\widetilde{K} = \left(\emptyset(X_i) - \frac{1}{N}\sum_{k=1}^{N}\phi(x_k)\right)'\left(\emptyset(X_j) - \frac{1}{N}\sum_{k=1}^{N}\phi(x_k)\right)$$

$$= K(X_i\,,X_j) - \frac{1}{N}\sum_{k=1}^{N}K(X_i\,,X_k) - \frac{1}{N}\sum_{k=1}^{N}K(X_j\,,X_k) + \frac{1}{N^2}\sum_{l,k=1}^{N}K(X_l\,,X_k)$$

$$\widetilde{K} = K - 2\left(1_{\frac{1}{N}}\right)K + \left(1_{\frac{1}{N}}\right)K\left(1_{\frac{1}{N}}\right) \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad \cdots\cdots\cdots\cdots (22)$$

حيث ان :

$\widetilde{K}$: تمثل مصفوفة اللب الجديدة بعد اجراء عملية التوسيط .

$\left(1_{\frac{1}{N}}\right)$: يمثل مصفوفة مربعة من درجة (MxM) جميع عناصرها تساوي $(\frac{1}{N})$.

4-3 طرائق تقدير عرض الحزمة(معلمة التمهيد) (h) :

bandwidth estimating Methods (smooting parameter) :

قبل التطرق الى الطرائق المستعملة في تقدير معلمة التمهيد لا بد من معرفة ماهية هذه المعلمة :

وهي عبــارة عن معلمــة حرة (Free Parameter)، و لها تأثير واضح في عملية التقدير لكونها تؤثر بشكل كبير في التحيز والتباين و بزيادة عرض الحزمة يزداد التحيز ويتناقص التباين والعكس صحيح ونتيجة لذلك تؤثر هذه المعلمة في درجة تمهيد المنحنى واقترابه من المنحنى الحقيقي .

لتقدير قيمة عرض الحزمة (معلمة التمهيد) (h) توجد العديد من الطرائق , سنركز في هذا البحث على مجموعة منها وكما يلي :

-

1-4-3 طريقة العبور الشرعي $(Cross - Validation)$:

تعد طريقة العبور الشرعي احدى الطرائق المعروفة لتقدير معلمة التمهيد (h) وتقسم الى :

1-1-4-3 طريقة المربعات الصغرى للعبور الشرعي ويرمز لها (LSCV) :

$(Least\ Squared\ Cross\ Validation)$:[17],[16],[3],[2],[1]

اقترحت طريقة المربعات الصغرى للعبور الشرعي في عام (1982/ م) من قبل (Rudemo) وتعد من افضل الطرائق وأكثرها استعمالاً لأيجاد قيمة معلمة التمهيد (h) من خلال تقدير (ISE) من البيانات المتوفرة وتصغيرها عند كل (h) .

$$ISE(h) = \int_{-\infty}^{\infty} \left[\hat{f}(x) - f(x)\right]^2 dx$$

$$= \int_{-\infty}^{\infty} \hat{f}^2(x)dx - 2\int_{-\infty}^{\infty} \hat{f}(x)f(x)dx + \int_{-\infty}^{\infty} f^2(x)dx \qquad \dots(23)$$

حيث نجد ان الحد الوحيد المراد تقديره في المعادلة اعلاه هو :

$$\int_{-\infty}^{\infty} \hat{f}(x)f(x)dx$$

لانه الحد الوحيد الذي يعتمد على معلمة التمهيد , أما بالنسبة للحدين الاخرين فأن :

$\int_{-\infty}^{\infty} f^2(x)dx$ مقدار لا يعتمد على قيمة معلمة التمهيد $(\ h\ )$.

$\int_{-\infty}^{\infty} \hat{f}^2(x)dx$ **قيمة معلومة تحسب من البيانات .**

وبطرح الحد الثابت نرى بأن تقليل (ISE) بالنسبة الى (h) يكون مساوي الى تقليل:

$$ISE(h) - \int_{-\infty}^{\infty} f^2(x)dx = \int_{-\infty}^{\infty} \hat{f}^2(x)dx - 2\int_{-\infty}^{\infty} \hat{f}(x)f(x)dx \qquad \dots(24)$$

و ان :

$$\int_{-\infty}^{\infty} \hat{f}(x)f(x)dx = \mathrm{E}\,\hat{f}(x) \qquad \dots(25)$$

ولتقدير $\mathrm{E}\,\hat{f}(x)$ في المعادلة (25) بأستعمال $[Leave - one - out\ Cross\ Validation]$ نحصل على :

$$\mathrm{E}\,\hat{f}(x) = n^{-1}\sum_{i=1}^{n} \hat{f}_{-i}(X_i) \qquad \dots(26)$$

وبتعويض المعادلة (26) في المعادلة (25) نحصل على $LSCV$ كالاتي:

$$LSCV(h) = \int_{-\infty}^{\infty} \hat{f}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(X_i) \qquad \dots(27)$$

علماً ان :

$$\hat{f}_{-i}(X_i) = (n-1)^{-1}h^{-1}\sum_{j\neq i}^{n} K\left(\frac{X_i - X_j}{h}\right) \qquad \dots(28)$$

وكذلك فأن :

$$\int_{-\infty}^{\infty} \hat{f}^2(x)dx = n^{-2}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} K*K\left(\frac{X_i - X_j}{h}\right) \qquad \dots(29)$$

و ان :

$K*K(z)$ **: يمثل الالتفاف لدالة Kernel**

$$z = \left(\frac{X_i - X_j}{h}\right)$$

ومن خلال تعويض المعادلة رقم (28) والمعادلة رقم (29) في المعادلة رقم (27) نحصل على :

$$\boldsymbol{LSCV(h) = n^{-2}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} K*K\left(\frac{X_i - X_j}{h}\right) - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}(X_i)}$$

$$\boldsymbol{= \frac{2}{n^2h}\left[\frac{n}{2}K*K(0)\right.}$$

$$\boldsymbol{\left. + \sum_{i=1}^{n-1}\sum_{j=1}^{n} K*K\left(\frac{X_i - X_j}{h}\right) - \frac{2n}{n-1}K\left(\frac{X_i - X_j}{h}\right)\right] \quad \ldots(30)}$$

علماً ان

$$\boldsymbol{K*K(0) = \int K(z)K(0-z)dz = \int K^2(z)dz = \|K\|_2^2 = C_K = R(K) \quad \ldots(31)}$$

وان المعلمة التمهيدية التي تعمل على تقليل الدالة :

$$\boldsymbol{\hat{h}_{LSCV} = argmin_h LSCV(h) \qquad \ldots(32)}$$

وفي عام (1987) أطلق الباحثان (Scott & Terrel) تسمية العبور الشرعي غير المتحيز على الدالـة $LSCV(h)$ لان :

$$\mathrm{E}(LSCV(h)) + \int_{-\infty}^{\infty} f^2(x)dx \qquad \ldots(33)$$

هو تقدير غير متحيز لـــ $MISE(\hat{f}(x))$

خوارزمية طريقة المربعات الصغرى للعبور الشرعي (LSCV)

1- افتراض قيم أولية للمعلمة التمهيدية (h) .

2- حذف المشاهدة $\{X_i\}$.

3- حساب مقدر (Leave-One-Out) $\hat{f}_{-i}(X_i)$ في المعادلة (2-55) عند نقاط المشاهدات.

4- تكوين دالة العبور الشرعي غير المتحيز.

$$LSCV(h) == n^{-2}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} K*K(\frac{X_j - X_i}{h}) - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}(X_i)$$

5- تكرر الخطوات المذكورة آنفا ،وفي كل مرة تحذف مشاهدة واحدة.

6- ثم اختيار المعلمة التمهيدية المقابلة لاصغر $LSCV(h)$ ،بحيث ان :

$$\hat{h}_{LSCV} = \arg \min_{h \in H_n} [LSCV(h)]$$

2-1-4-3 طريقـة العبـور الشرعي المتحيـز ($Biased\ Cross\ Validation$) ويرمز لها (BCV) [12],[7],[6]],[4],[1]:

اقترحت طريقة العبور الشرعي المتحيز في عام (1987/ م) من قبل (Terrell و Scott) وتعتمد هذه الطريقة على استعمال (AMISE) بدلاً من (MISE) حيث ان :

$$\mathbf{AMISE(h) = (nh)^{-1}R(K) + h^4 \left(\frac{M_2(K)}{2}\right)^2 R(f^2)}$$

$$\mathbf{= \frac{R(K)}{nh} + \frac{h^4 d_v^2}{4} R\left(\acute{\acute{f}}\right)} \quad \dots (34)$$

اذ ان :

$$\mathbf{M_2(K)} = \int_{-\infty}^{\infty} \boldsymbol{x^2\, g(x) dx}$$

$$R(K) = \int_{-\infty}^{\infty} K^2(z) dz$$

وان :

$$\boldsymbol{R\left(\acute{\acute{f}}\right) = R(f^2) = \int_{-\infty}^{\infty} \left[\acute{\acute{f}}(x)\right]^2 dx} \quad \dots (35)$$

وللحصول على دالة $BCV(h)$ يتم استبدال القيم المجهولة في $R\left(\acute{\acute{f}}\right)$ بالمقدار الاتي :

$$\boldsymbol{\tilde{R}\left(\acute{\acute{f}}\right) = R\left(\hat{\acute{\acute{f}}}(.,h)\right) - \frac{R\left(\acute{\acute{K}}\right)}{nh^5} = n^{-2} \sum\sum_{i \neq j} \left(\acute{\acute{K}}_h * \acute{\acute{K}}_h\right)(X_i - X_j)} \quad \dots (36)$$

ولاثبات الصيغة اعلاه:

$$\boldsymbol{R\left(\hat{\acute{\acute{f}}}\right) = \frac{1}{n} \acute{\acute{K}}_h * \acute{\acute{K}}_h(0) + \frac{1}{n^2} \sum\sum_{i \neq j} \acute{\acute{K}}_h * \acute{\acute{K}}_h(Xi - Xj)}$$

$$\boldsymbol{= \frac{1}{nh^5} \acute{\acute{K}} * \acute{\acute{K}}(0) + \frac{1}{n^2 h^5} \sum\sum_{i \neq j} \acute{\acute{K}} * \acute{\acute{K}}(Xi - Xj)}$$

$$= \frac{1}{nh^5} R\left(\acute{\acute{K}}\right) + \frac{1}{n^2h^5} \sum\sum_{i\neq j} \acute{\acute{K}} * \acute{\acute{K}}(Xi - Xj) \qquad \dots (37)$$

اذا كانت h مثلى ومن الدرجة $\left(n^{-4/5}\right)$ فأن $R\left(\acute{\acute{\hat{f}}}\right)$ هي تقدير متحيز لـ $R\left(\acute{\acute{f}}\right)$ حيث:

$$E\left(R\left(\acute{\acute{\hat{f}}}\right)\right) = R\left(\acute{\acute{f}}\right) + \frac{R\left(\acute{\acute{K}}\right)}{nh^5} + o(h^2) \qquad \dots (38)$$

ونتيجة لذلك اقترحوا لتقدير $R\left(\acute{\acute{f}}\right)$ يستخدم $\widetilde{R}\left(\acute{\acute{f}}\right)$ وبذلك

$$\widetilde{R}\left(\acute{\acute{f}}\right) = R\left(\acute{\acute{\hat{f}}}\right) - \frac{R\left(\acute{\acute{K}}\right)}{nh^5} \qquad \dots (39)$$

والتي تقود الى:

$$BCV(h) = \frac{R(K)}{nh} + \frac{h^4 d_{Kv}^2}{4n^2} \sum\sum_{i\neq j} \acute{\acute{K}}_h * \acute{\acute{K}}_h\left(X_i - X_j\right) \qquad \dots (40)$$

وان المعلمة التمهيدية التي تعمل على تقليل الدالة:

$$\hat{h}_{BCV} = \operatorname{argmin}_{h\in H_n}[BCV(h)] \qquad \dots (41)$$

خوارزمية طريقة (BCV):

1. نفترض القيمة الاولية لمعلمة التمهيد (h) .
2. نعوض قيم كل من $R(K)$ و $M_v^2 = d_v^2$ لدالة (Gaussian) من الجدول رقم (3-2) ادناه .

جدول رقم (3-2)

يمثل قيم $R(K), M_v^2(K), C_v(K)$ لدالة (Gaussian)

Kernel	v	R(K)	$M_v^2(K)$	$C_v(K)$
Gaussian	2	$\frac{1}{2\sqrt{\pi}}$	1	1.06

3. نحسب قيمة المقدار : $\sum\sum_{i\neq j} \acute{\acute{K}}_h * \acute{\acute{K}}_h(Xi - Xj)$ وفق الصيغة ادناه:

$$\sum\sum_{i\neq j} \acute{\acute{K}}_h * \acute{\acute{K}}_h(Xi - Xj) = \sum\sum_{i\neq j} \varphi^{\acute{\acute{}}}_{\sqrt{2h^2}}(Xi - Xj) \qquad \dots (42)$$

علماً ان: $\varphi_{\sqrt{2h^2}}^{''}$ **تمثل المشتقة الثانية لدالة كثافة (Gaussian) اما** $\sqrt{2h^2}$ **يمثل الانحراف المعياري لها وكذلك نجد قيمة** $R\left(\acute{\acute{K}}\right)$**.**

4. **نجد قيمة** $R\left(\acute{\acute{\hat{f}}}\right)$ **من المعادلة (40).**
5. **نجد قيمة** $\widetilde{R}\left(\acute{\acute{f}}\right)$ **والتي تظهر صيغتها بالمعادلة (42).**
6. **نحسب** $BCV(h)$ **الظاهرة في المعادلة (43).**
7. **نختار المعلمة التمهيدية التي تعمل على تقليل الدالة:**

$$\hat{\mathbf{h}}_{\mathbf{BCV}} = \mathbf{argmin}_{\mathbf{h}\in\mathbf{H_n}}[\mathbf{BCV}(\mathbf{h})]$$

3-1-4-3 طريقة العبـور الشـرعي الممهـد ويرمز لها (SCV)
$(Smoothed\ Cross-Validation)$: [18],[16],[6],[4],[1]

اقترحت طريقة للعبور الشرعي الممهد في عام (1992/ م) من قبل (Park and Marron , Hall) والتي تعتمد على $MISE(h)$ **:**

$$MISE(h) = IV(h) + IB(h) \qquad \ldots(43)$$

اذ ان :

$$IV(h) = \int_{-\infty}^{\infty} var\left(\hat{f}(x)\right)dx = \frac{C_K}{nh} + \frac{1}{n}\int_{-\infty}^{\infty}(K_h * f)^2(x)dx \qquad \ldots(44)$$

$$IB(h) = \int_{-\infty}^{\infty} bias^2\left(\hat{f}(x)\right)dx = \int_{-\infty}^{\infty}(K_h * f - f)^2(x)dx \qquad \ldots(45)$$

وبتعويض المعادلة رقم (2-76) والمعادلة رقم (2-77) في المعادلة رقم (2-75) تصبح :

$$= \int_{-\infty}^{\infty}(K_h * f)^2(x)dx - 2\int_{-\infty}^{\infty}(K_h * f)(x)f(x)dx + \int_{-\infty}^{\infty} f^2(x)dx \qquad \ldots(46)$$

الإ انه في عام (1992/ م) اقترح كل من $(Marron\ and\ Wand)$ **:**

- **بأن** $\left(\frac{R_K}{nh}\right)$ **يمثل تقدير جيد لـ** $IV(h)$

- وفي نفس العام استعمل مقترحي هذه الطريقة مصطلح $I\widehat{V}(h)$.
- أما بخصوص تقدير الـ $IB(h)$ اقترحوا ان يقدروا f بـ $\hat{f}_g(x)$ علماً بأن $\hat{f}_g(x)$ مقدر ثاني.

$$I\widehat{B}(h) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(K_h * K_h - 2K_h + K_0) * L_g * L_g(X_i - X_j) \qquad \dots(47)$$

اذ ان :

g : تمثل المعلمة التمهيدية او ما تسمى عرض الحزمة التجريبية .

L : تمثل دالة **Kernel**

وهما يختلفان عن المعلمة التمهيدية h ودالة K .

K_0 : تعرف بدالة $(Dirac)$

وبأستعمال التقريب $n \approx n-1$ اشتقوا الدالة :

$$SCV(h) = \frac{C_K}{nh} + \frac{1}{n(n-1)}\sum\sum_{i \neq j}(K_h * K_h - 2K_h + K_0) * L_g * L_g(X_i - X_j)$$

$\dots(48)$

وان المعلمة التمهيدية التي تعمل على تقليل الدالة :

$$\hat{h}_{SCV} = \mathrm{argmin}_{h \in H_n}[SCV(h)] \qquad \dots(49)$$

خوارزمية طريقة العبور الشرعي الممهد (SCV):

1. نفترض القيمة الاولية لمعلمة التمهيد h
2. نحسب قيمة $C_K = R(K)$ والتي تساوي وحسب دالة (Gaussian) $\frac{1}{2\sqrt{\pi}}$
3. نحسب قيمة $I\widehat{V}(h)$ والتي تساوي $(\frac{C_K}{nh})$.
4. نحسب قيمة تقدير $I\widehat{B}(h)$ وفق الصيغة التالية:

$$I\widehat{B}(h) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j \neq i}^{n}\left(\varphi_{\sqrt{2h^2+2g^2}} - 2\varphi_{\sqrt{h^2+2g^2}} + \varphi_{\sqrt{2g^2}}\right)(X_i - X_j)$$

$\dots(50)$

5. نحسب قيمة

$$\hat{g} = \left(\frac{15}{16\sqrt{\pi}\,v_6}\right)^{1/7} n^{-1/7} \qquad \dots(51)$$

علماً ان قيمة v_6 تحسب كالاتي:

$$v_6 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{1}{h^6} \left[\frac{b_{ij}^6}{h^6} - 15\frac{b_{ij}^4}{h^4} + 45\frac{b_{ij}^2}{h^2} - 15 \right] \varphi_h(b_{ij}) \quad \dots (52)$$

حيث ان :

$$b_{ij} = (X_i - X_j) \quad \dots (53)$$

6. نختار المعلمة التمهيدية التي تعمل على تقليل الدالة:

$$\hat{h}_{SCV} = \mathrm{argmin}_{h \in H_n} [SCV(h)]$$

<u>3-4-2 طريقة قاعدة الملئ المباشر $(Direct\ Plug-in\ Rule)$</u> [1],[2],[3],[13],[14],[17]

اقترحت هذه الطريقة من $(Sheather\ and\ Thompson\ , Scott)$ وتعتمد هذه الطريقة على تعويض تقديرات الكميات المجهولة في صيغة المعلمة التمهيدية المثلى بدلالة قيم Ψ_r، و ان :

$$R(f^v) = \int_{-\infty}^{\infty} [f^v(x)]^2 dx \quad \dots (54)$$

v : تمثل درجة المشتقة.

وكون $R(f^v)$ كمية مجهولة لذلك سنستعمل الصيغة ادناه لتقديرها :

$$\Psi_r = \int f^r(x) f(x) dx$$

$$\Psi_r = E(f^r(x)) \quad \dots (55)$$

علماً ان ($r = 2v$) .

وبالتالي نحصل على :

$$R(f^v) = (-1)^v \Psi_{2v} = (-1)^{(r/2)} \Psi_r \quad \dots (56)$$

وبالاعتماد على عرض الحزمة التجريبية (g) فأن مقدر Ψ_r هو :

$$\hat{\Psi}_r(g) = n^{-1} \sum_{i=1}^{n} \hat{f}^r(Xi, g)$$

$$= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} L_g^r(Xi - Xj) \qquad ...(57)$$

علماً ان

$$\hat{f}^r(x,g) = n^{-1}g^{-r-1}\sum_{j=1}^{n} L^r\left(\frac{x - Xj}{g}\right) \qquad ...(58)$$

وبتعويض المعادلة رقم **(58)** في المعادلة رقم **(57)** نحصل على :

$$\hat{\Psi}_r(g) = \frac{1}{\mathrm{n(n-1)}g^{r+1}}\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{j=1}}^{\mathrm{n}} L^r\left(\frac{Xi - Xj}{g}\right) \qquad ...(59)$$

وان ايجاد $\hat{\Psi}_r(g)$ تعتمد على اختيار عرض الحزمة التجريبي (g)، ولايجاد عرض الحزمة الامثل والناتج من تقليل $(AMSE)$ نحتاج اولاً لحساب $\hat{\Psi}_r^{NS}$ وفق المعادلة التالية:

$$\hat{\Psi}_r^{NS} = \frac{(-1)^{\frac{r}{2}}\ r!}{(2\,\sigma)^{r+1}\ \left(\frac{r}{2}\right)!\ \sqrt{\pi}} \qquad ...(60)$$

اذ سيتم الاعتماد على هذا المقدر لايجاد القيمة الاولى للمعلمة التجريبية وهذا ما سيتم توضيحه في خوارزمية هذه الطريقة، اما المعلمة التجريبية فيتم حسابها وفق الصيغة التالية:

$$g_{AMSE} = \left[\frac{-k!\ L^r(0)}{M_k(L)\ \Psi_{\mathrm{r}+k}\ \mathrm{n}}\right]^{1/(r+k+1)} \qquad ...(61)$$

علماً بأن دالة **(Kernel)** دالة متماثلة وان $(k = 2, 4, \ldots)$ وتمتلك r من المشتقات.

وبالاعتماد على المعلمة التمهيدية (g) نحصل على قاعدة الملئ المباشرة DPI اي $\hat{h}_{DPI}$

$$\hat{h}_{DPI} = \left[\frac{R(K)}{M_2^2\ (K)\ \hat{\Psi}_{r+k}(g)\ n}\right]^{1/(r+k+1)} \qquad ...(62)$$

خوارزمية طريقة (DPI):

الخطوات المتبعة لمرحلتين $(l = 2)$ ومعنى ذلك نحتاج لايجاد (g_2, g_1) كما يلي وبأستعمال دالة (Gaussian)

1. نحسب قيمة $\widehat{\Psi}_8^{NS}$ وفق قاعدة التوزيع الطبيعي وحسب المعادلة $(h = \hat{\sigma}\,(1.06)\,n^{-1/5})$ وتبلغ قيمتها:

$$\widehat{\Psi}_8^{NS} = \frac{105}{32\sqrt{\pi}\,(\sigma)^9} \quad \dots (63)$$

2. نحسب قيمة g_1 وفق المعادلة (61) وكما يلي:

$$g_1 = \left[\frac{-2!\ \ L^6(0)}{M_2(L)\ \ \widehat{\Psi}_8^{NS}\ \ n}\right]^{1/9} \quad \dots (64)$$

3. نحسب قيمة $\widehat{\Psi}_6(g_1)$ وفق المعادلة (58) وكما يلي:

$$\widehat{\Psi}_6(g_1) = \frac{1}{n(n-1)\ g_1^7}\sum_{i=1}^{n}\sum_{j=1}^{n} L^6\left(\frac{X_i - X_j}{g_1}\right) \quad \dots (65)$$

4. نحسب قيمة g_2 وكما يلي :

$$g_2 = \left[\frac{-2!\ \ L^4(0)}{M_2(L)\ \widehat{\Psi}_6(g_1)\ \ n}\right]^{1/7} \quad \dots (66)$$

5. نحسب قيمة $\widehat{\Psi}_4(g_2)$ وكما يلي:

$$\widehat{\Psi}_4(g_2) = \frac{1}{n(n-1)\ g_2^5}\sum_{i=1}^{n}\sum_{j=1}^{n} L^4\left(\frac{X_i - X_j}{g_2}\right) \quad \dots (67)$$

6. نجد قيمة المعلمة التمهيدية حسب طريقة (DPI) وفق المعادلة (62)

$$\hat{h}_{DPI} = \left[\frac{R(K)}{M_2^2(K)\ \ \widehat{\Psi}_4(g_2)\ \ n}\right]^{1/5} \quad \dots (68)$$

وفق دالة (Gaussian) يتم حساب ما يلي:

$$L^4(u) = \frac{u^4}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} - \frac{6\,u^2}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} + \frac{3}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$$

$$L^6(u) = \frac{u^6}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} - \frac{15\,u^4}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} + \frac{45\,u^2}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2} - \frac{15}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$$

$$L^4(0) = \frac{3}{\sqrt{2\pi}}$$

$$L^6(0) = -\frac{15}{\sqrt{2\pi}}$$

ونعوض قيم كل من $R(K)$ و $M_2^2(K)$ من الجدول رقم (3-2) .

4 - الجانب التطبيقي :

4-1المقدمة: (introduction)

لغرض تطبيق ما ورد في الجانب النظري وبما ينسجم وهدف الدراسة ولتحقيق ذلك الهدف جرى استعمال اسلوب المحاكاة بغية محاكاة عدد كبير جدا من الحالات المقترحة التي يمكن ظهورها ضمن الواقع العملي وبذلك يمكن للنتائج ان تكون اكثر شمولية , تم اعداد برنامج بلغة (R) وتنفيذ تجارب المحاكاة بأستعمال ثلاث حجوم للعينات , تمثلت العينة الاولى بأنها ذات الحجم الصغير (n=20) , وتم اختيار حجم العينة الثانية المتوسطة (n=50) , أما العينة الثالثة اقترحت ذات حجم كبير (n=100) , وبتكرار (Replicatec=1000) لكل عينة , أما بالنسبة لعدد المتغيرات فكان (p=12) متغير .

وكانت النتائج كما يلي :

جدول رقم (4-1) يمثل نتائج الجذور المميزة وما يقابلها من نسبة تفسير للتباين الكلي لدالة (RBF KERNEL) عند استخدام طرق مختلفة لتقدير عرض الحزمة (h) وبحجم عينة (n=20)

	eig.scv	var.scv	مجموع التباين المفسر	eig.lscv	var.lscv	مجموع التباين المفسر	eig.pi	var.pi	مجموع التباين المفسر	eig.bcv	var.bcv	مجموع التباين المفسر
Comp.1	1.585066	12.12960	12.12960	1.946285	15.2864	15.2864	1.7682406	13.21091	13.21091	1.554551	11.97380	11.97380
Comp.2	1.24518	9.55371	21.6833	1.4279096	11.27033	26.5567	1.3318857	9.97751	23.18842	1.2269996	9.46901	21.44281
Comp.3	1.118156	8.59038	30.2737	1.233285	9.73938	36.2961	1.171618	8.78846	31.97688	1.1044	8.53023	29.97304
Comp.4	1.05322	8.10046	38.3742	1.117152	8.81969	45.1158	1.084817	8.14670	40.12358	1.0445034	8.07215	38.04519
Comp.5	1.020288	7.85260	46.2268	1.03533	8.15158	53.2674	1.0350731	7.77944	47.90302	1.0154665	7.84980	45.89499
Comp.6	1.006053	7.74526	53.972	0.98654	7.74499		1.0115482	7.60541	55.50843	1.0041827	7.76296	53.65795
Comp.7	1.00124	7.70899	61.681	0.954224	7.45829		1.0023072	7.53712	63.04555	1.0007251	7.73625	61.3942
Comp.8	1.000069	7.70013	69.3811	0.922487	7.16817		0.999304	7.51480		1.0000643	7.73124	69.12544
Comp.9	0.9995	7.69588		0.882169	6.81415		0.9964628	7.49394		0.9999959	7.73060	
Comp.10	0.998063	7.68529		0.829209	6.36981		0.9895179	7.44316		0.9998503	7.72951	
Comp.11	0.993783	7.65369		0.766487	5.85585		0.9748984	7.33655		0.9989127	7.72247	
Comp.12	0.984405	7.58402		0.699755	5.32117		0.9516038	7.16642		0.994966	7.69230	

جدول رقم (4-2) يمثل نتائج الجذور المميزة وما يقابلها من نسبة تفسير للتباين الكلي لدالة (RBF KERNEL) عند استخدام طرق مختلفة لتقدير عرض الحزمة (h)

جدول رقم (4-2) يمثل نتائج الجذور المميزة وما يقابلها من نسبة تفسير للتباين الكلي لدالة (RBF KERNEL) عند استخدام طرق مختلفة لتقدير عرض الحزمة (h)
وبحجم عينة (n=50)

	eig.scv	var.scv	مجموع التباين المفسر	eig.lscv	var.lscv	مجموع التباين المفسر	eig.pi	var.pi	مجموع التباين المفسر	eig.bcv	var.bcv	مجموع التباين المفسر
Comp.1	1.5576839	19.64780	19.64780	1.7489089	20.18879	20.18879	1.7434919	20.6689	20.6689	1.5182011	19.47260	19.47260
Comp.2	0.9149174	11.57258		1.0500756	12.23954	32.4283	0.9935941	11.80775		0.8968569	11.52013	
Comp.3	0.7731424	9.78854		0.8994376	10.49857		0.8362109	9.94300		0.7584013	9.74791	
Comp.4	0.6827503	8.65066		0.7947034	9.27963		0.7340450	8.73403		0.6707902	8.62547	
Comp.5	0.6166684	7.81915		0.7070844	8.26360		0.6575068	7.82801		0.6075572	7.81449	
Comp.6	0.5655123	7.17668		0.6330972	7.39748		0.5967587	7.11037		0.5587233	7.18905	
Comp.7	0.5242895	6.65944		0.5707794	6.66922		0.5492984	6.54961		0.5193574	6.68429	
Comp.8	0.4910344	6.24211		0.5165128	6.03373		0.5083394	6.06646		0.4877576	6.28019	
Comp.9	0.4661698	5.92995		0.4688604	5.47745		0.4762259	5.68802		0.4639387	5.97457	
Comp.10	0.4463918	5.68200		0.4297942	5.01395		0.4523482	5.40697		0.4454784	5.73858	
Comp.11	0.4310105	5.48877		0.3995758	4.64837		0.4331427	5.18034		0.4305787	5.54764	
Comp.12	0.4193395	5.34192		0.3710819	4.29016		0.4193518	5.01684		0.4194714	5.40519	

جدول رقم (4-2) يمثل نتائج الجذور المميزة وما يقابلها من نسبة تفسير للتباين الكلي لدالة (RBF KERNEL) عند استخدام طرق مختلفة لتقدير عرض الحزمة (h) وبحجم عينة (n=100)

	eig.scv	var.scv	مجموع التباين المفسر	eig.lscv	var.lscv	مجموع التباين المفسر	eig.pi	var.pi	مجموع التباين المفسر	eig.bcv	var.bcv	مجموع التباين المفسر
Comp.1	1.7023206	23.77177	23.77177	1.3825315	20.23449	20.23449	1.8470237	24.17745	24.17745	1.6622383	23.66774	23.66774
Comp.2	0.8359809	11.67751		0.7910016	11.91420		0.8991700	11.76982		0.8183000	11.65376	
Comp.3	0.7171193	10.01739		0.6978439	10.53609		0.7718376	10.10254		0.7017406	9.99434	
Comp.4	0.6338775	8.85689		0.6266595	9.49069		0.6824345	8.93317		0.6202669	8.83657	
Comp.5	0.5686033	7.94489		0.5685219	8.62358		0.6116658	8.00674		0.5562146	7.92415	
Comp.6	0.5115390	7.15118		0.5136512	7.80416		0.5490374	7.18838		0.5007880	7.13715	
Comp.7	0.4614492	6.45454		0.4622447	7.04087		0.4936474	6.46560		0.4518608	6.44237	
Comp.8	0.4149675	5.80853		0.4123219	6.29920		0.4401589	5.76837		0.4079316	5.81834	
Comp.9	0.3723183	5.21616		0.3640184	5.57566		0.3911115	5.13012		0.3671239	5.23897	
Comp.10	0.3362266	4.71431		0.3154725	4.84817		0.3464317	4.54716		0.3335260	4.76038	
Comp.11	0.3081624	4.32285		0.2682596	4.12976		0.3119593	4.09723		0.3077058	4.39208	
Comp.12	0.2896091	4.06369		0.2290593	3.50306		0.2902973	3.81398		0.2896320	4.13474	

- مناقشة النتائج والتوصيات:

1-5 مناقشة النتائج :

من خلال النتائج المبينة في الجداول (4-1) و (4-2) و (4-3) اعلاه نجد ان استخدام دالة (RBF KERNEL) في تحليل المركبات الرئيسية اعطت أفضل النتائج في حالة حجم العينة الصغيرة (n=20) عند تقدير معلمة التمهيد (عرض الحزمة) بطريقة العبور الشرعي الممهد (SCV) لتفسيرها اعلى نسبة من التباين الكلي ضمن المركبات الرئيسية الفعالة (التي تزيد القيمة الذاتية لها عن الواحد الصحيح) تليها طريقة العبور الشرعي المتحيز(BCV) ومن ثم التقدير بطريقة قاعدة الملئ المباشر (PI) وجاءت طريقة المربعات الصغرى للعبور الشرعي (LSCV) في المرتبة الاخيرة .

أما في حالة حجم العينة المتوسط (n=50) فكانت اعلى نسبة مفسرة من التباين الكلي ضمن المركبات الرئيسية الفعالة عند استخدام طريقة المربعات الصغرى للعبور الشرعي (LSCV) تليها طريقة قاعدة الملئ المباشر (PI) وجاءت في المرتبة الثالثة طريقة العبور الشرعي الممهد (SCV) وأخيراً طريقة العبور الشرعي المتحيز (BCV) لتفسيرها أقل نسبة من التباين الكلي .

وبالنسبة للحالة الثالثة وهي حجم العينة الكبير (n=100) سجلت النتائج اعلى نسبة مفسرة من التباين الكلي ضمن المركبات الرئيسية الفعالة عند تقدير معلمة التمهيد بطريقة قاعدة الملئ المباشر (PI) بالمرتبة الاولى وتليها طريقة العبور الشرعي الممهد (SCV) بالمرتبة الثانية وجاءت طريقة العبور الشرعي المتحيز (BCV) بالمرتبة الثالثة تليها طريقة المربعات الصغرى للعبور الشرعي (LSCV) بالمرتبة الأخيرة .

أما بالنسبة لتقدير معلمة عرض الحزمة (h) (معلمة التمهيد) , فأن افضل طريقة لتقدير معلمة التمهيد عند حجم عينة (n=20) كانت طريقة العبور الشرعي الممهد هي الافضل في تقدير معلمة التمهيد , وعند حجم عينة (n=50) كانت طريقة المربعات الصغرى للعبور الشرعي هي الافضل , وعند حجم عينة (n=100) كانت طريقة قاعدة الملئ المباشر هي الافضل .

2-5 التوصيات :

من خلال النتائج التي تم التوصل اليها والاستنتاجات والملاحظات السابقة , فأن من أهم التوصيات التي يوصي بها الباحث هي :

1. **اسخدام الدوال اللبية في تحليل المركبات الرئيسية (KPCA) في حالة مجاميع البيانات متعددة المتغيرات الغير خطية .**
2. **الدقة في اختيار معلمة التمهيد (عرض الحزمة (h)) والطريقة الملائمة لتقديها لما لها من تأثير واضح في عملية ايجاد المركبات الرئيسية الفعالة التي تزيد قيم الجذور الذاتية لها عن الواحد الصحيح وما يقابلها من النسبة في تفسير التباين الكلي**

3- التوسع في استخدام الدوال اللبية في الظواهر الاقتصادية والسلوكية والانسانية لما تتمتع به من مرونة كبيرة في التعامل مع الظواهر المدروسة .

المصادر

1. ابراهيم , مروة خليل (2013م) " تقويم بيانات العمر والجنس للتعدادات السكانية مع تطبيق عملي لبيانات التعداد العام للسكان لسنة 1997 في العراق" رسالة ماجستير / كلية الادارة والاقتصاد / جامعة بغداد .
2. حمود، مناف يوسف (2005م) "مقارنة المقدرات اللامعلمية لتقدير دوال الكثافة الاحتمالية" اطروحة دكتورا / كلية الإدارة والاقتصاد / جامعة بغداد.
3. *Abdel-Razzaq and Jetter, J. (2010) "A SIMULATION STUDY FOR THE BANDWIDTH SELECTION IN THE KERNEL DENSITY ESTIMATION BASED ON THE EXACT AND THE ASYMPTOTIC MISE" Pak. J. Statist., Vol. 26 ,pp. 239-265.*
4. *Chouaib, C.; Mohamed-Faouzi, H.; Messaoud, D. (2015) " Adaptive kernel principal component analysis for nonlinear dynamic process monitoring" Appl. Math. Inf. Sci. 9, No. 4, pp. 1833-1845 .*
5. *Duong, T. (2004) "Bandwidth Selectors for Multivariate Kernel Density Estimation" Ph.D. dissertation, School of Mathematics and Statistics, University of Western, Australia.*
6. *Duong, T. and Hazelton, M. L. (2005) "Cross-validation Bandwidth Matrices for Multivariate Kernel Density Estimation" Board of the Foundation of the Scandinavian Journal of Statistics, Blackwell Publishing Ltd, Vol. 32, pp. 485–506.*
7. *Guidoum , A (2014) " Kernel Estimator and Bandwidth Selection for Density and its Derivatives" The kedd Package in R.*
8. *Hal Daume (2012) " A Course in Machine Learning" Data Science 101.*
9. *Jong-Min Lee, ChangKyoo Yoo, Sang Wook Choi, Peter A. Vanrolleghem, In-Beum Lee (2004) "Nonlinear process monitoring using kernel principal component analysis" Chemical Engineering Science , voi . 59 , pp .223 – 234 .*
10. *Kenji Fukumizu (2012) " Kernel Methods for Statistical Learning" Machine Learning Summer School 2012, Kyoto , http://www.ism.ac.jp/~fukumizu/MLSS2012/.*

11. *Lecture about (2007) "Nonparametric Density and Regression Estimation", http://athena.sas.upenn.edu/petra/class222/nonpar.pdf .*
12. *Molanes-López, E. M. and Cao, R. (2008) "Plug-in bandwidth selector for the kernel relative density estimator" Annals of the Institute of Statistical Mathematics AISM (60(2)), pp. 273-300.*
13. *Raykar V. C. and Duraiswami R. (2005) "Very fast optimal bandwidth selection for univariate kernel density estimation" Department of Computer Science and Institute for Advanced Computer Studies University of Maryland, CollegePark.*
14. *Schölkopf, B; Smola, A; Müller, K (1998) "Nonlinear Component Analysis as a Kernel Eigenvalue Problem" IEEE Journals & Magazines , Vol . 10, Issue: 5 , pp . 1299 – 1319 .*
15. *Sain, S. R., Baggerly, K. A., Scott, D. W. (1992) "Cross-Validation of Multivariate Densities" Office of Naval Research.*
16. *Tenreiro, C. (2011) "Combining cross-validation and plug-in methods for kernel density bandwidth selection" CMUC and DMUC, University of Coimbra.*
17. *Turlach, B. A. (1993) "Bandwidth Selection in Kernel Density Estimation: A Review" C.O.R.E. and Institut de Statistique. Universit e Catholique de Louvain.*