بحوث مستقبلية (١٨) ٢٠٠٧
ص ص [١٨٥-١٩٧]

تقدير الملاءمة الجيدة لمجموعة نقاط معطيات باعتماد الإجراءين Polyfit و Polyval

محمد جاجان يونس الزبيدي[١]
لمى أكرم عبدالله الصفـــار[٢]

الملخص

أكدت المَعلَمات أن ظهور الخط المستقيم يكون على شكل مسار لنقاط كثيرة تقع باتجاه محدد إلا أنها تتباين في كثافتها من موقع لآخر حسب التشوهات التي تصيب تلك النقاط والذي يؤدي إلى عدم ظهور الشكل الدقيق لقطعة الخط المستقيم. تم في هذا البحث اعتماد علاقة خطية مفترضة من الدرجة الأولى تُعطي تطابقاً أكثر للبيانات وذلك بتطبيق لغة الـMatlab v.7.2 في تحديد الملاءمة اللازمة للخط المرسوم بالتقدير الخطي مع قيم نقاط معطيات باستخدام الإجراءين (Polyfit & Polyval)، إذ يتم تقليل الخطأ بين الدالة البسيطة وقيم البيانات وذلك بإيجاد الدالة التقريبية التي يكون فيها مجموع مربعات قيم الانحرافات للفرق بين الدالة ونقاط المعطيات الحقيقية أقل ما يمكن.

Abstract

The parameters confirmed that the appearance of the straight line is being as a path of many points which placed in a certain direction. However these points differ in its consistency from one place to another depending on the distortion which infects these points, which lead to the fact that the straight line does not appear in its accurate shape. In this paper, we have used the virtual linear relation of first order degree which give more identical data by using Matlab language version 7.2 to defined the best line points fit estimating using Polyfit and polyval procedures while reducing the error between the simple function and given line points by finding the estimating function which the summation of derivation line points for difference between the function and the input line points is minimum.

(١) مدرس مساعد، قسم علم الحاسوب، كلية الحدباء الجامعة.
(٢) مدرس مساعد، قسم علوم الحاسبات، كلية التربية، جامعة الموصل.

تاريخ الاستـــلام: ٢٠٠٦/١٠/١٠
تاريخ قبول النشر: ٢٠٠٦/١١/٠٧

المقدمة(Introduction):

تبقى مسألة استخلاص الخطوط المستقيمة من المواضيع المهمة جداً في معالجات الصور الرقمية وتمييز الأنماط، وذلك لكون الخوارزميات المتوافرة قد عالجت إمكانية الاكتشاف من دون تحديد بداية أو نهاية(استمرارية) الخط المستقيم، وكذلك عدم معالجة النقاط التي تحيد بعض الشيء عن مسارها كخط مستقيم إذ عدّتها نقاطاً لا تنتمي إلى الخط المستقيم ولا تساهم بتكوينه[Robert and Michael, 1997]. وقد عالجت الكثير من البحوث عمليات تحسين الصورة باستخدام أساليب كثيرة كعمليات بسط(مط) المستويات الرمادية للصورة (Contrast Stretch Enhancement) أو من خلال استخدام المرشحات(Filters) التي تعمل على التنقية من دون معالجة التشوهات التي تعرضت لها الصورة بوصف تلك التشوهات من مكونات الصورة ذاتها، إذ يلاحظ وبشكل واضح التعرجات التي تحدث لنقاط الخطوط المستقيمة والتي تؤدي إلى فقدان المعالم الحقيقية لتلك الخطوط[الفهادي والسيف، ٢٠٠٢]. تم في هذا البحث اعتماد أسلوب يعتمد على معادلات الانحدار الخطي البسيط من الدرجة الأولى والذي يُعد أسلوباً مناسباً في تقدير أفضل خط مستقيم لنقاط(بيانات) معطاة من خلال إيجاد دالة بسيطة تناسب المدى الكامل لتلك النقاط، وهذا ما يُعرف بملاءمة الخطوط (Line Fitting)، [DuMouchel and O'Brien, 1989].

الانحدار الخطي(Linear Regression):

يطلق الإحصائيون اسم الانحدار الخطي على عملية تحديد المعادلة الخطية والذي يعد الملاءمة الجيدة لنقاط معطاة، إذ يتم تحديد المعادلة الخطية بتقليص مجموع مربع المسافات بين الخط ومجموعة نقاط معطيات وذلك من خلال تقليل الخطأ بين الدالة البسيطة وقيم النقاط بإيجاد الدالة التقريبية التي يكون فيها مجموع مربعات قيم الانحرافات للفرق بين الدالة والبيانات الحقيقية أقل ما يمكن، ويُعرف هذا الأسلوب بطريقة ملاءمة المنحني باستخدام طريقة المربعات الصغرى(Least Square Curve Method)، [سيفي وكمال الدين، ١٩٨٦]، إذ يمكن تمثيل التعبير الرياضي للخط المستقيم من خلال المعادلة(١)، [Levenberg, 1944]:

$$y = b_0 + b_1 x + e \qquad ....(1)$$

حيث: b_0, b_1 معاملان يمثلان المقطع الصادي والميل على التوالي.

x, y متغيران يمثلان إحداثيات أيّة نقطة على الخط المستقيم.

e تمثل الخطأ.

وأما النقاط الموزعة حول الخط المستقيم ذات الإحداثيات$(x_i\ ,\ y_i)$ فيمكن تحديد مواقعها باستخدام المعادلة(٢):

$$y = \hat{b}_o + \hat{b}_1 x \qquad ....(2)$$

حيث: $\hat{b}_o , \hat{b}_1$ يمثلان أفضل معاملات مقدّرة للخط المستقيم المار ضمن النقاط المعتمدة.

ولتحديد الخط المستقيم يمكن اعتماد حساب التقديرات للمعاملات$(b_o\ , b_1)$ باستخدام طريقة المربعات الصغرى، إذ أن:

$$\hat{b}_o = \bar{y} - \hat{b}_1 \bar{x} \qquad ....(3)$$

$$\hat{b}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad ....(4)$$

حيث: $\bar{x} , \bar{y}$ يمثلان الوسط الحسابي للقيم x , y على التوالي.

n تمثل عدد نقاط الخط المستقيم.

وبذلك تكون معادلة الخط المستقيم كالآتي:

$$y = b_o + b_1 x \qquad ....(5)$$

ثم يتم رسم الخط المستقيم من النقاط التي تحقق المعادلة(5) والتي تعطي تطابقاً أكثر لنقاط الخط المستقيم اعتماداً على العلاقة الخطية المفترضة كما في المعادلة(6,7)، [Kovesi, 2004]:

$$b_0 = \frac{\sum y_i (\sum x_i)^2 - \sum x_i \sum x_i y_i}{m(\sum x_i)^2 - (\sum x_i)^2} \qquad ....(6)$$

$$b_1 = \frac{m \sum x_i y_i - \sum x_i \sum y_i}{m(\sum x_i)^2 - (\sum x_i)^2} \qquad ....(7)$$

حيث: m تمثل عدد نقاط الخط المستقيم.

b_0 , b_1 معاملان يمثلان المقطع الصادي والميل على التوالي.

بعد ذلك يتم تعويض المعاملين$(b_0\ , b_1)$ المقدّرتين في المعادلة(5) لرسم الخط المستقيم وإعطاء أكثر ملاءمة لنقاط المعطيات.

ولتوضيح كيفية الحصول على خط مستقيم للوصل بين مجموعة نقاط نفترض وجود مجموعة من نقاط المعطيات التي تم جمعها من نتائج تجربة عملية، يتم أخذ هذه القيم ورسمها في مستوي إحداثي، مما يظهرها على شكل نقاط موزعة على المستوي الإحداثي ويبدو جلياً أن هذه النقاط لا تقع على خط مستقيم يصل بينها، وإذا تم رسم خط مستقيم فإنه لن يمر إلا بنقطتين من مجموعة نقاط المعطيات المتوفرة، عليه يتم الحصول على خط مستقيم أقرب ما يكون إلى مجموعة النقاط المرسومة في المستوي الإحداثي على وفق طريقة ملاءمة المنحني باستخدام طريقة المربعات الصغرى وذلك من خلال تقليص مربع المسافة بين كل نقطة من مجموعة النقاط والخط المستقيم المرسوم.

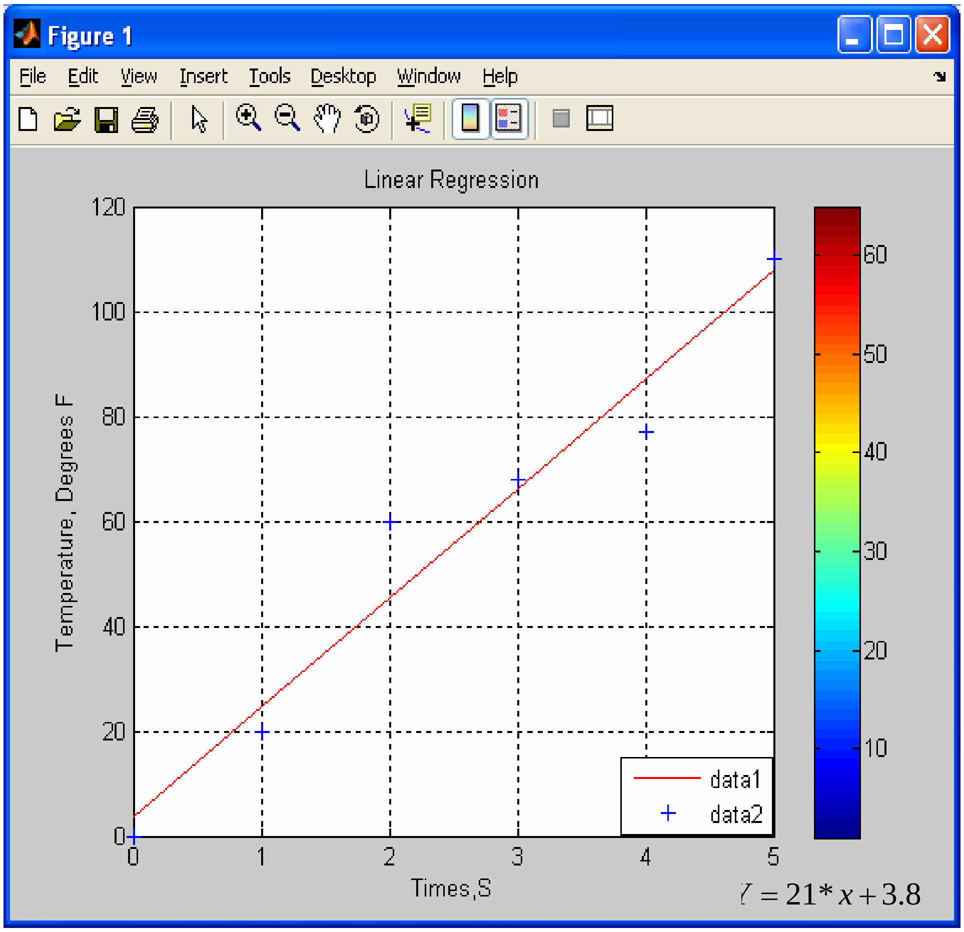
ولفهم تأثير الانحدار الخطي بشكل عملي يمكن عدّ وجود مجموعة من القيم التي تعبر عن تغير درجات الحرارة بوحدة الفهرنهايت مع الزمن بوحدة الثانية لمدة زمنية قدرها خمس ثوانٍ ثم توضع القيم في الجدول(١) وعلى وفق الشروط الآتية:

١. قيم المحور الأفقي هي(٠–٥).
٢. قيم المحور العمودي هي(٠–١٢٠).
٣. رسم منحني الارتداد الخطي(Linear Regression).
٤. رسم مجموعة نقاط المعطيات في المستوي الإحداثي(x, y).
٥. تسمية المحور الأفقي بمحور الزمن(Time) بوحدة هي الثانية(S).
٦. تسمية المحور العمودي بمحور درجة الحرارة(Temperature) بوحدة هي الدرجة فهرنهايت(Degree F).
٧. تسمية الشكل بالانحدار الخطي(Linear Regression).

الجدول(١) قيم تمثل تغير درجة الحرارة مع الزمن[عارف ونعمة، ٢٠٠٣].

Time, S.	Temperature, F.
0.0	0.0
1.0	20.0
2.0	60.0
3.0	68.0
4.0	77.0
5.0	110.0

بعد رسم نقاط الجدول أعلاه في مستوي إحداثي باستخدام التقدير الخطي نلاحظ أن هذه النقاط تكون قريبة لترسم خطاً مستقيماً من خلال إظهار مجموعة قيم النقاط المعطاة، الشكل(١).



الشكل(١) ملاءمة أفضل خط مستقيم لبيانات الجدول(١).

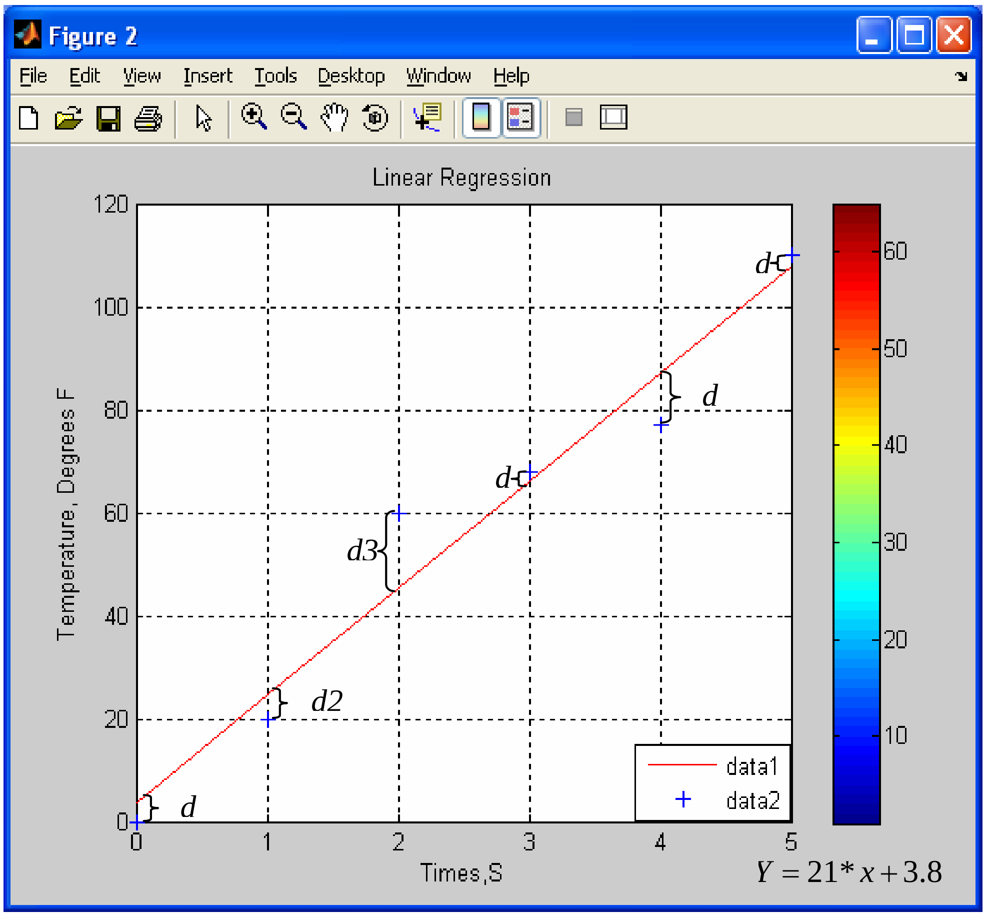
ولتحديد كمية الملاءمة اللازمة للخط المرسوم بالتقدير الخطي مع قيم نقاط المعطيات، سوف يتم تحديد المسافة بين كل نقطة من نقاط المعطيات مع الخط المرسوم، الشكل(٢) وحسب المعادلة(8)، [Daniel and Wood, 1980]:

$$d = y - \hat{y} \qquad \text{...................(8)}$$

حيث: y تمثل قيمة أو نقطة مجدولة.

$\hat{y}$ تمثل مقدّر النقطة المجدولة.

d تمثل الانحراف.



الشكل(٢) انحرافات نقاط معطيات الجدول(١).

لو فرضنا وجود(m) من القيم المجدولة($x_1, x_2, x_3,........., x_m$) وتقابلها(m) من القيم ($y_1, y_2, y_3,.........y_m$) فتكون الانحرافات كما في المعادلة(9)، [Arbenz, 2005]:

$$\left.\begin{aligned} d_1 &= y_1 - \hat{y}_1 = y_1 - f(x_1) \\ d_2 &= y_2 - \hat{y}_2 = y_2 - f(x_2) \\ d_3 &= y_3 - \hat{y}_3 = y_3 - f(x_3) \\ &\vdots \\ d_m &= y_m - \hat{y}_m = y_m - f(x_m) \end{aligned}\right\} \qquad ....(9)$$

من المعادلة(9) يلاحظ أن بعض قيم الانحرافات تكون موجبة والأخرى سالبة فيكون الناتج الكلي صفراً نسبة إلى خط الملاءمة الجيد، ولتجاوز الإشارات السالبة يتم حساب مجموع قيم الانحرافات($\sum(md_i)$) من المعادلة(10):

$$s = \sum_{i=1}^{m} d_i^{\ 2} = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \qquad ....(10)$$

حيث: s تمثل مجموع مربعات قيم الانحرافات.

الانحدار لمعادلة متعدد الحدود(Polynomial Regression):

في الفقرة السابقة تم توضيح كيفية حساب الملاءمة الجيدة لمعادلة خطية معبرة عن مجموعة نقاط معطيات، وأما في هذه الفقرة فسيتم شرح كيفية استعمال متعدد حدود واحد لملاءمة المعطيات من خلال تقليص المسافة الفاصلة بين متعدد الحدود ومجموعة نقاط معطيات.

بدايةً فإن المعادلة الخطية هي(متعدد حدود) من الدرجة الأولى، وأما صيغة المعادلة العامة لمتعدد الحدود ذو متغير واحد فهي كالآتي[عارف ونعمة، ٢٠٠٣]:

$$f(x) = a_0 x^N + a_1^{\ N-1} + a_2^{\ N-2} + ... + a_{N-1}x + a_N \qquad ....(11)$$

حيث: N تمثل مجموعة من النقاط تُستخدم لتحديد متعدد حدود من الدرجة N .

إن الخط المرسوم سابقاً على وفق التقدير الخطي يطلق عليه الرياضيين في حالتنا هذه أسم ملاءمة متعدد الحدود الخطية(Polynomial Linear Fitting) من الدرجة الأولى. ويتم على وفق لغة الـMatlab حساب قيم المقطع الصادي والميل الخاصين لحساب المجموع الحسابي لمربع المسافة الفاصلة بين النقاط والخط المرسوم على وفق ملاءمة متعدد الحدود من الدرجة الأولى وذلك باستخدام الإجراءين(Polyfit&Polyval) لحساب الملاءمة الجيدة لمجموعة نقاط معطيات.

الإجراء(Polyfit):

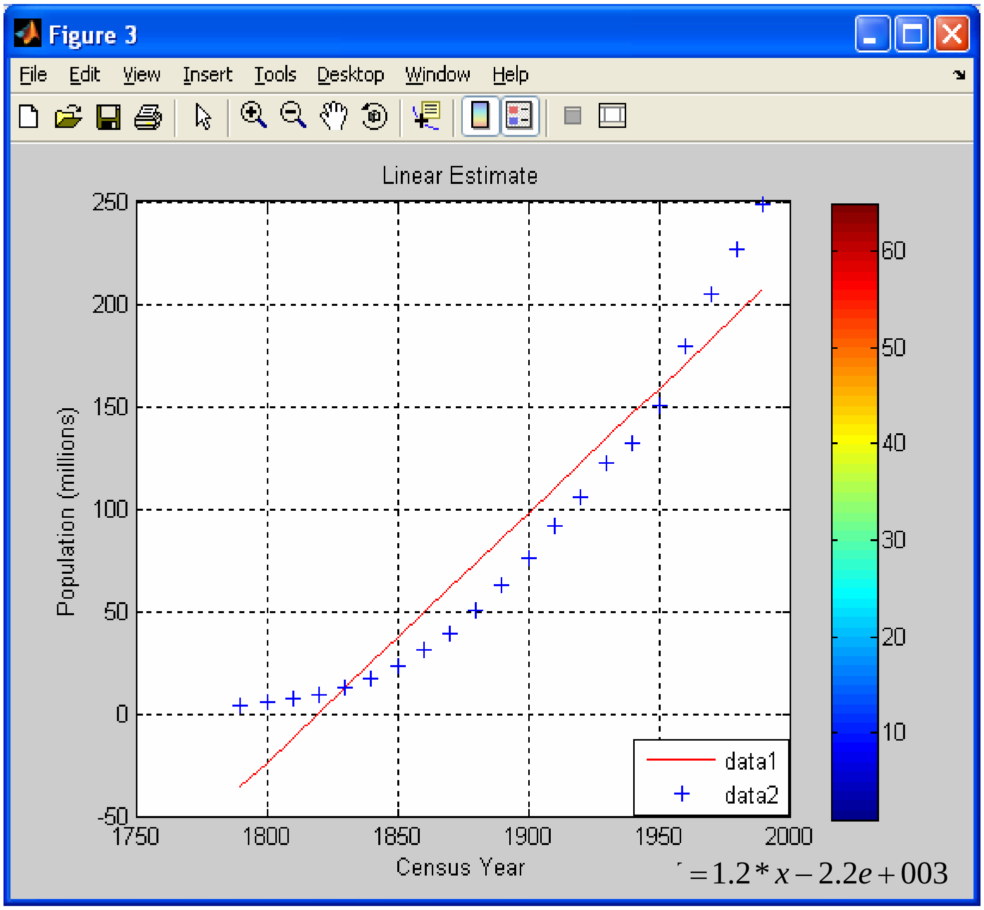
من المعادلة(3) أعلاه تبين أنه كلما ازدادت درجة متعدد الحدود ازدادت معه عدد النقاط التي يمر بها الخط المستقيم، ونتيجة لذلك نحصل على أن كل النقاط(N) ستقع على متعدد الحدود والذي سيسبب تغيرات كبيرة بين نقاط المعطيات، وهذا يقود إلى أخطاء في تقدير قيم الفترات بشكل كبير، عليه تم استخدام لغة الـMatlab لحساب الملاءمة الجيدة لمجموعة نقاط معطيات باستخدام متعدد حدود من خلال تعليمة الـ(Polyfit) إذ تملك هذه التعليمة ثلاثة متغيرات(X,Y,N) يُعبِّر المتغيران(X,Y) عن المستوي الإحداثي لمجموعة نقاط المعطيات، بينما يُعبِّر المتغير(N) عن درجة متعدد الحدود المستخدم.

الإجراء(Polyval):

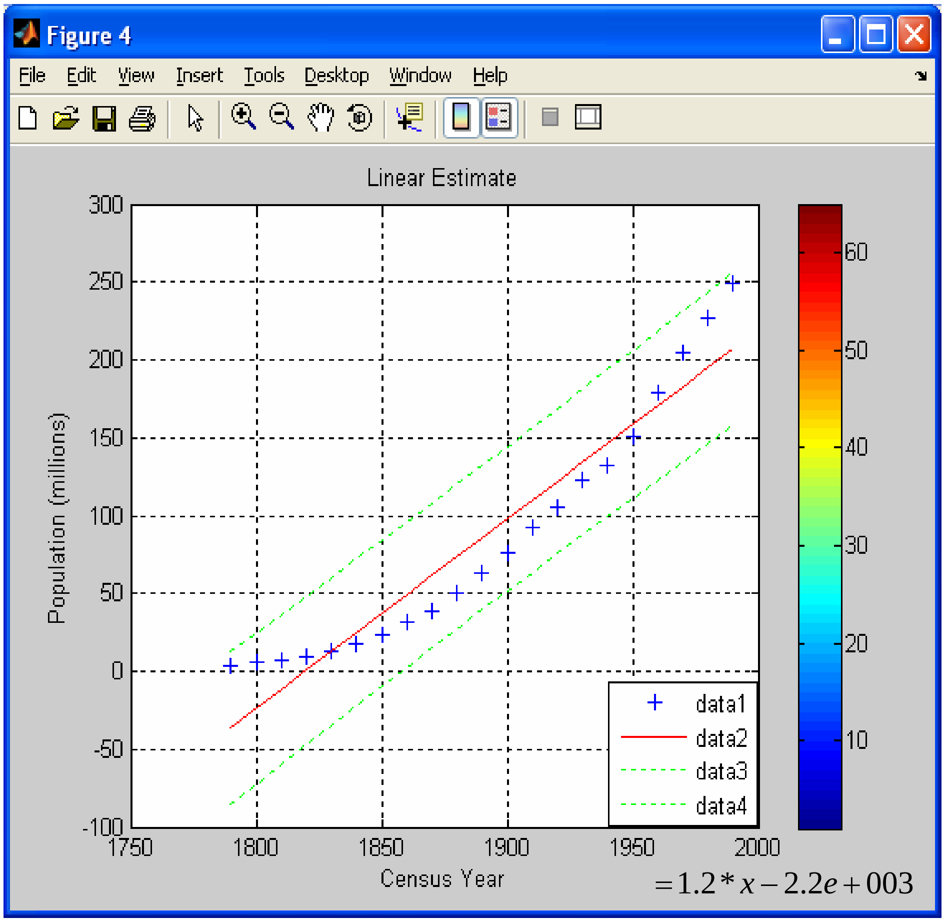
يتم تطبيق هذه التعليمة لأجل حساب نقاط منحني الانحدار الخطي من خلال قيم المعاملات، إذ تملك هذه التعليمة متغيرين يدخلان في تشكيل التعليمة، المتغير الأول هو معاملات متعدد الحدود، بينما يمثل المتغير الثاني قيم الشعاع(X) المعبرة عن القيم المستخدمة من متعدد الحدود والمحددة من قبل المبرمج في لغة الـMatlab.

وكمثال تطبيقي آخر تم من خلال اللغة البرمجية المستخدمة اعتماد معطيات(Census) والتي تمثل إحصائية لسكان الولايات المتحدة الأميركية منذ سنة(1790) كأقل قيمة(min) ولغاية سنة(1990) كأعلى قيمة(max) بمعدل(mean=1890) وإنحراف معياري (std=62.0484) للمدى السكاني(range=200)، [Whaley and Jack, 2003]، الشكل(٣). بينما يمثل الشكل(٤) تحديداً لحدود الأخطاء(Error Bounds) للملاءمة اللازمة للخط المرسوم بالتقدير الخطي مع قيم نقاط معطيات(Census) على وفق الشروط الآتية:

١. قيم المحور الأفقي هي(١٧٥٠–٢٠٠٠).
٢. قيم المحور العمودي هي(–٥٠–٢٥٠).
٣. رسم منحني الارتداد الخطي(Linear Regression).
٤. رسم مجموعة نقاط المعطيات في المستوي الإحداثي(x,y).
٥. تسمية المحور الأفقي(Census Year) بوحدة هي السنة(Year).
٦. تسمية المحور العمودي بمحور السكان(Population) بوحدة هي المليون(Millions).
٧. تسمية الشكل بالتقدير الخطي(Linear Estimation).



الشكل(٣): ملاءمة أفضل خط مستقيم لبيانات(Census).



الشكل(٤): حدود الأخطاء في ملاءمة أفضل خط مستقيم لبيانات(Census).

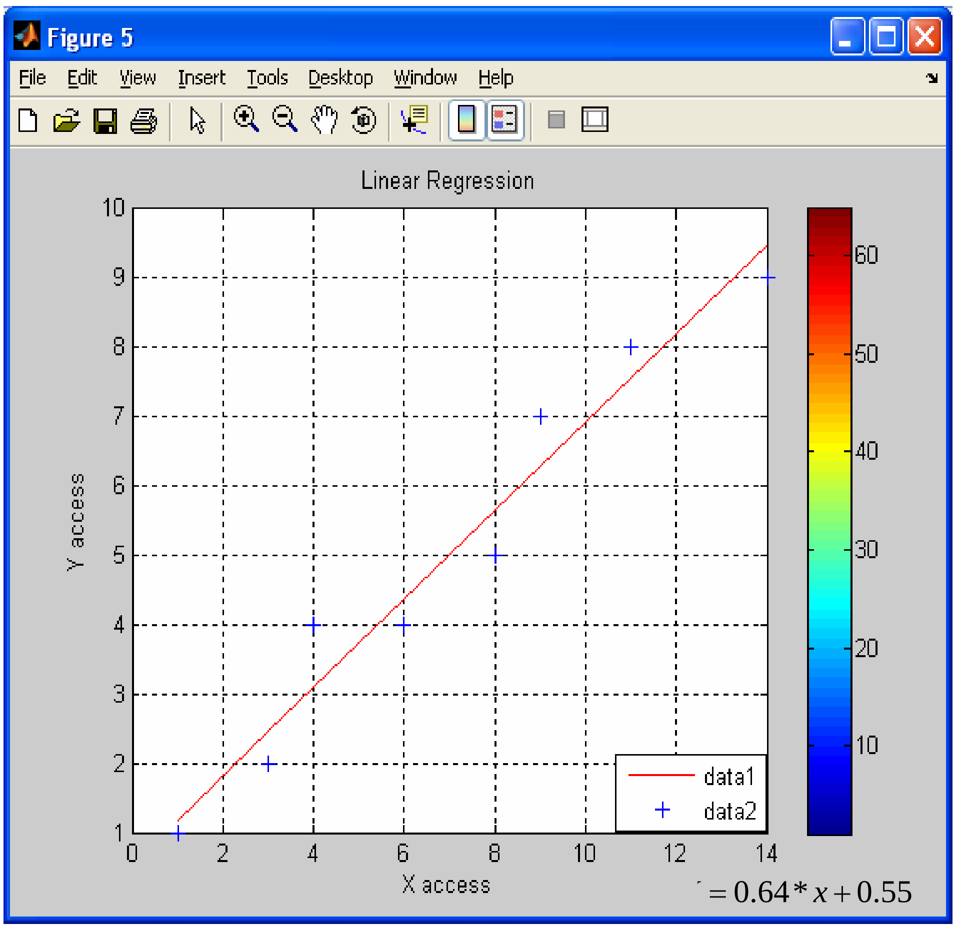
ولتوضيح آلية العمل في الأمثلة أعلاه تم افتراض قيم تمثل بيانات تحيد عن مسارها كخط مستقيم المتمثلة بالجدول(٢) وبتطبيق معادلات الارتداد الخطي على إحداثيات تلك البيانات وباستخدام التعليمتين(Polyfit, Polyval) على افتراض دالة خطية بصيغة الخط المستقيم كما في المعادلة(1) ومن ثم حساب النقطتين (b_0 , b_1) ثم إيجاد مجموع القيم وتعويضها في المعادلتين(6,7) لإعطاء تطابق أكثر لعدد(m) من القيم، الشكل(٥) وكما يأتي:

الجدول(٢): بيانات تحيد عن مسارها كخط مستقيم.

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

$$b_0 = \frac{40*524-56*364}{8*524-(56)^2} = \frac{6}{11} = 0.545 \approx 0.55$$

$$b_1 = \frac{8*364-56*40}{8*524-(56)^2} = \frac{7}{11} = 0.636 \approx 0.64$$



الشكل(٥): ملاءمة أفضل خط مستقيم لبيانات الجدول(٢).

النتائج(Results):

تم في هذا البحث عرض بعض الأمثلة التطبيقية لقيم تمثل مجموعة نقاط معطيات تحيد عن مسارها كخط مستقيم. وقد أظهر التطبيق العملي نتائج تعد إيجابية، ويؤمل أن تؤدي البحوث الجارية إلى تحسين هذه النسبة. ولتسهيل تنفيذ البرنامج المقترح واستغلال كفاءة البيئة البرمجية للحاسبة تم استخدام لغة Matlab v.7.2 لما تتمتع به من مواصفات فنية عالية في مجالات الاستخدام المختلفة ولما لها من مميزات السرعة في التنفيذ والإمكانية في استخدام الدوال الجاهزة والقدرة على توليد أعداد كبيرة من البيانات.

الاستنتاجات(Conclusions):

تمت ملاحظة الاستنتاجات الآتية التي ظهرت من خلال النتائج التطبيقية للدراسة المقترحة والتي يمكن أن تؤدي إلى تكامل نسبي للعاملين في هذا المجال:

إن الطريقة المتبعة لملاءمة مجموعة نقاط معطيات باستخدام متعدد حدود من الدرجة الأولى أظهرت ملاءمة جيدة لخطوط منتظمة ذات استقامة عالية عند تطبيقها عملياً على نماذج جداول مختلفة تمثل نقاطاً متفرقة تحيد عن مسارها كخط مستقيم.

تطبيق مفاهيم الانحدار الخطي في ملاءمة نقاط معطاة والحصول على معادلة الخط المستقيم أدى إلى ملاءمة أفضل خط مستقيم يمثل تلك النقاط ومتغيراً باتجاهها العام وحدود مسارها.

إن افتراض دالة خطية بصيغة الخط المستقيم لإحداثيات نقاط مختارة في لغة الـMatlab وباستخدام الإجراءين(Polyfit&Polyval) أدى إلى الحصول على خط مستقيم بأقل انحراف ممكن.

من الممكن أن لا تقع أي من مجموعة النقاط المعطيات على الخط الناتج بالملاءمة الجيدة (Best Fit) لمجموعة النقاط المرسومة في المستوي الإحداثي بالتقدير الخطي.

المصادر (References):

١. **سيفي، علي محمد صادق وكمال الدين، ابتسام كمال، (١٩٨٦): "مبادئ التحليل العددي"، وزارة التعليم العالي والبحث العلمي، مديرية دار الكتب للطباعة والنشر، ص ص ١٦٥-١٧٦، جامعة بغداد، بغداد، العراق.**

٢. **عارف، محمد ونعمة، مازن(٢٠٠٣): "المرجع الأساسي لتعلم MATLAB 6.X "، البراق للطباعة والنشر والتوزيع، الطبعة الأولى، حلب، سورية.**

٣. **الفهادي، قبيس سعيد والسيف، خليل إبراهيم(٢٠٠٢): "تحسين صورة الخط المستقيم في الصور الثنائية باعتماد الانحدار الخطي"، المجلة العراقية للعلوم الإحصائية، العدد(٣)، المجلد(٢)، ص(١-١٠)، كلية علوم الحاسبات والرياضيات، جامعة الموصل، الموصل، العراق.**

4. Arbenz P.(2005): "Fitting Lines, Rectangles and Squares in the Plane", Institute of Computational Science, Zurich, Switzerland.
5. Daniel, C. and Wood, F.S.(1980): "Fitting Equations to Data", John Wiley & Sons, New York, USA.
6. Draper, N.R and Smith H.(1998): Applied Regression Analysis, 3rd Ed., John Wiley & Sons, New York, USA.
7. DuMouchel, W. and O'Brien, F.(1989): "Integrating a Robust Option into a Multiple Regression Computing Environment", in Computing Science and Statistics: Proceedings of the 21st Symposium on the Interface, American Statistical Association, Alexandria, VA, pp. 297-301, USA.
8. Kovensi P.(2004): "Least Squares Fit Line to Set of Points", School of Computer Science and Software Engineering, University of Western Australia.
9. Levenberg, K.(1944): " Method for the Solution of Certain Problems in Least Squares", Quart. Appl. Math, Vol. 2, pp. 164-168, USA.
10. Robert A.M. and Michael D.A.(1997): "Hough Transform", the Technical Report-Hough Transform,
11. Whaley R. C. and Jack D.(2003): "Curve Fitting Toolbox for Use With Matlab", the Mathworks Inc., MA, USA.