

نماذج متعدد المتغيرات باستعمال الشبكات العصبية الاحصائية

مروان عبد الحميد عاشور

جامعة بغداد - بغداد - جمهورية العراق

dr_Marwan2012@yahoo.com

المستخلص:

ان الغرض من هذه الورقة هو دراسة معنوية تاثير متغيرات المستقلة في نماذج متعدد المتغيرات باستعمال نماذج ذكية تتصف بالمرونة وحادثة ممثلاً بالشبكات العصبية الاحصائية العصبية لمحاصيل الزراعية الاستراتيجية التي تؤثر على الامن الغذائي للدول. واهم ماتوصل اليه من خلال هذا البحث هو اعتماد الشبكات العصبية الاحصائية بشكل خاص والشبكات العصبية الاصطناعية بشكل عام كبدل عن الطرائق التقليدية المعتمدة لمعالجة نماذج متعدد المتغيرات كالتحليل القويم المعتمدة وخصوصاً في العلاقات او النماذج او شبه خطية او غير خطية او حتى الخطية ، مما يوفر اداة قوية ومرنة لمعالجة نماذج متعدد المتغيرات تتصف بالكفاءة والكفاءة التامة.

الكلمات المفتاحية: شبكة الانحدار العام، نماذج متعدد المتغيرات، شبكة العصبية الاحصائية.

Multivariate Models using Statistical Neural Networks

Marwan Abdul Hammed Ashour

University of Baghdad – Baghdad-Iraq

dr_Marwan2012@yahoo.com

Abstracts:

The purpose of this paper is to study the effect of independent variables in multivariate models using intelligent models characterized by flexibility and modernity represented by Neural Networks for strategic agricultural crops that affect the countries' nutritional security. The most important findings of this research are the adoption of statistical neural networks in particular and artificial neural networks in general as an alternative to the traditional methods adopted to deal with multivariate models such as correct analysis, especially in relationships, models, quasi-linear, nonlinear or even linear, Flexible and smart to handle multivariate models with complete efficiency.

Keyword: General Regression Network, Multivariate models, Networks Statistical Neural.

المقدمة:

تعد الشبكات العصبية الاصطناعية من أحدث وسائل معالجة البيانات في العالم ومازالت البحوث مستمرة في هذا المجال للتعرف على كفاءة هذه الوسائل. يهدف البحث الى الإفادة من خاصية عملية تعلم وتدريب ذاتيا مع أي نموذج التي تمتلكها الشبكات العصبية الاصطناعية دون الحاجة الى بناء أنموذج رياضي الذي تتطلبه الطرائق التقليدية وفي بعض الأحيان تكون عملية النمذجة معقدة وصعبة حسب طبيعة الظاهرة أو بيانات مما يتطلب مهارات رياضية كبيرة فضلاً عن جهدا ووقت. وتم انتخاب بيانات معتمدة في دراسات وبحوث سابقة لبيان كفاءة وأمثلي الشبكات العصبية. ان عينة البحث تشمل مجموعتين من البيانات حيث اعتمدت كمية انتاج المحاصيل الزراعية الاستراتيجية كمتغيرات للمجموعة الاولى تمثل " كمية انتاج الحنطة، كمية انتاج الشعير، كمية انتاج الرز" على التوالي اما متغيرات المجموعة الثانية فتتمثل " المتوسط العام للمساحة المزروعة (حنطة، شعير، رز) المعدل العام لسعر (الحنطة، الشعير، الرز)، كمية المياه السطحية، معدل النمو السنوي للسكان" اما الحد الزمني للبحث فيمثل الفترة 1978-2006. تتبع اهمية البحث من خلال توظيف طرائق حديثة لتشخيص العوامل المؤثرة على انتاج المحاصيل الزراعية الاستراتيجية لغرض مساعدة متأخذ القرار من بناء سياسات وخطط انتاجية استراتيجية لهذه المحاصيل المهمة وتبرز اهمية الشبكات العصبية الاصطناعية في معالجة نموذج متعدد المتغيرات الى سهولة معالجة النماذج غير الخطية وشبه الخطية ولاتحتاج الى بناء نموذج رياضي الذي تتطلبه النماذج التقليدية الذي يكون معقد في بعض الحالات ويحتاج الى مهارات رياضية وخبرة كبيرة. اما منهجية البحث فنكمن في توظيف المنهج الاحصائي التحليلي المتقدم مرتكزا على تطبيقات الذكاء الاصطناعي في معالجة نماذج متعدد المتغيرات. تنقسم هيكلية البحث الى ثلاثة اجزاء رئيسية هي على التوالي الجانبي النظري والتطبيقي واهم الاستنتاجات التي تم التوصل اليها من هذه الورقة.

الجانب النظري:**متعدد المتغيرات:**

متعدد المتغيرات هو علاقة (خطية او غير خطية) بين واحد او اكثر (Y's) من المتغيرات المعتمدة او الاستجابة وواحد او اكثر (X's) من المتغيرات المستقلة او المتنبئ بها. نستعمل النموذج الخطي او غير الخطي لتمثيل علاقة بين y's و x's ومن ثم تقدير واختبار المعلمات في النموذج. واحدة من مزايا، اختيار المتغيرات التي يمكن تضمينها في النموذج اذا كان لا يعرف مسبقاً. [1,2,5]

نموذج متعدد المتغيرات:

يشير نموذج متعدد المتغيرات الى انه مجموعة من المتغيرات المعتمدة التي تتعلق بمجموعة من المتغيرات المستقلة. وفي هذه الحالة يتم قياس عدة y's المناظرة لكل مجموعة من x's. لكل (y_1, y_2, \dots, y_p) يكون متوقع او متنبأ به من قبل كل من (x_1, x_2, \dots, x_q) . والصيغة الاتية

$$\begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \\ \vdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} \end{pmatrix} \dots \dots (1)$$

اما فرضيات التي نقودنا الى تقديرات جيدة هي كالاتي:

1. $E(Y)=XB$ or $E(\varepsilon) = 0$
2. $Cov(y_i)=\sum$ for $i=1,2,\dots,n$, where y is the i th row of y
3. $cov(y_i, y_j)=0$ for all $i \neq j$

حيث ان:

$$cov(y_i) = \sum = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \dots \dots (2)$$

المقياس الاكثر استخداماً وعلى نطاق واسع للاقتران بين مجموعتين من البيانات هو الارتباط القويم.

ويمثل معامل الافاضة (Rd) نسبة التباين الحاصل في متغيرات مجموعة والمفسرة بمتغيرات مجموعة اخرى، قيمته تتراوح بين (0 و 1) ويحسب Rd الكلي لبيان اثر متغيرات المجموعة X's في متغيرات المجموعة Y's من خلال الصيغة الرياضية الاتية:

$$\bar{R}dy/x = \sum_1^p Rdy_t \dots \dots (3)$$

الشبكات العصبية الاحصائية: Statistical Neural Networks

يستخدم الاحصاء كذلك في الشبكات التغذية الامامية وواحد من اكثر اهمية لتلك الشبكات هو شبكة دالة الاساس الشعاعي. اصحبت الشبكات العصبية تنتشر بشكل متزايد بتطبيقات متنوعة وربما يكون المنافس الرئيسي لمستقبلات متعدد الطبقات. [4-6]

شبكة دالة الاساس الشعاعي: Radial Basis Function Networks

تعد اساليب تصنيف الانماط الاحصائية التقليدية الاساس او الملهم لشبكات دالة الاساس الشعاعي والتي يرمز لها اختصاراً بـ RBF، والتي تعد اساساً للحصول على زخماً جديداً

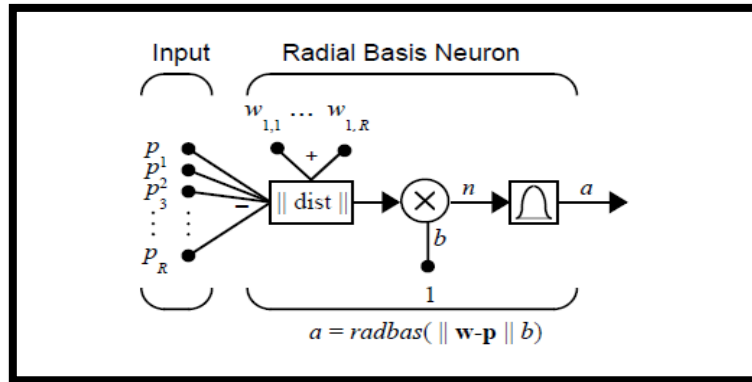
لوصفها الشبكات العصبية. وتتكون المعمارية الهندسية الاساسية لشبكات دالة الاساس الشعاعي RBF من ثلاث طبقات هي كالآتي: [5,6,8]

طبقة المدخلات : هي الطبقة الاولى وتمثل طبقة مدخلات الشبكة وهي لا تقوم بأي معالجة

الطبقة المخفية: هي الطبقة الثانية وتقوم بمعالجة العلاقات غير الخطية من فضاء المدخلات الى فضاء اعلى الابعاد والتي تكون قابلة للفصل خطياً.

طبقة المخرجات: هي الطبقة الاخيرة والتي تقوم بجمع الازان البسيطة لمخرجات الخطية.

اذا استخدمت شبكة RBF لتقريب دالة فإنها تعطي مخرجات جيدة، ومع ذلك اذا كان مطلوباً تصنيف نمط معين بدالة سينية Sigmoid function، فان معالجة مخرجات الشبكة يكون بعطاء قيمة للمخرج 0 او 1. والشكل الاتي يبين البنية الهندسية لشبكة دالة الاساس الشعاعي RBF.



شكل (1) يمثل البنية الهندسية لشبكة دالة الاساس الشعاعي RBF

هنالك عدة فوائد لشبكة دالة الاساس الشعاعي RBF مقارنة بمستقبلات متعدد الطبقات (MLP) منها ان تدريب شبكة RBF اسرع من MLP وانه يقدم حد قرار افضل والفائدة الاخرى تتمثل بان الطبقة المخفية في RBF اسهل تمثيل من الطبقة المخفية في MLP. وتمثل الطبقة المخفية في RBF بدالة كثافة الاحتمالية. وهناك بعض العيوب تتجلى في ان MLP تعطي تمثيل توزيع بشكل افضل من RBF.

شبكة الانحدار العام العصبية: General Regression Neural Network

تعد شبكة الانحدار العام العصبية من الشبكات التغذوية الامامية التي طورت من قبل Donald Specht عام 1991 لنمذجة نظام والتشخيص ويمكن اعتبارها كحالة عامة لشبكة العصبية الاحتمالية. ويرمز لها بـ GRNN. [7,9]

ان الشبكة الاحتمالية تستخدم لتصنيف الانماط خصوصاً، بينما شبكة الانحدار العام GRNN لها تطبيقات واسعة. تستند فكرة بناء شبكة الانحدار العام GRNN على امكانية تقريب أي

دالة تعطي مجموعة مدخلات وزوج بيانات من المخرجات. بافتراض ان الدالة لديها n مدخلات (x_1, \dots, x_n) ومخرج واحد (y) ، وقيمة المتوسط المتوقع لمخرج المعطى العائد لمدخل x ، يمكن ايجادها باستخدام المعادلة الاتية من خلال نظرية الاحتمالات. [5,8]

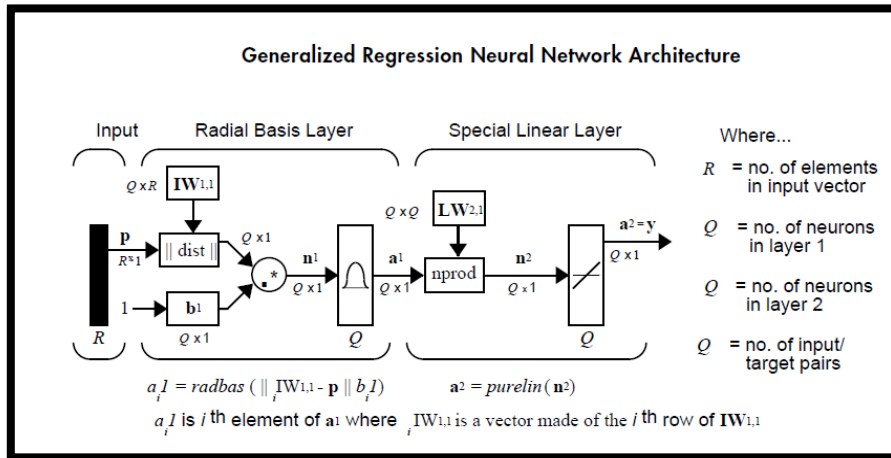
$$\bar{y} = \frac{\int_{-\infty}^{\infty} y pdf(x,y) dy}{pdf(x)} \dots \dots \dots (4)$$

هذه المعادلة تتضمن دالة كثافة الاحتمالية $pdf(x)$ التي هي دالة الكثافة الاحتمالية المشتركة. وهذا يعني ان احتمالية المخرج تصبح y واحتمالية المدخل تصبح x . تقرب دوال كثافة الاحتمال بواسطة جمع دالة كاوسين، لذا يمكن تقريب شرط دوال الاحتمالية ايضاً. ويمكن تقريب المعادلة اعلاه بواسطة المعادلة الاتية: [3,8]

$$\bar{y} = \frac{\sum_{p=1}^P Y_p e^{-d^2 / 2\sigma^2}}{\sum_{p=1}^P e^{-d^2 / 2\sigma^2}} \dots \dots \dots (5)$$

نفترض هذه المعادلة وجود P انواع لبيانات في مجموعة التدريب. تمثيل معمارية شبكة الانحدار العام العصبية GRNN من خلال الشكل الاتي:

شكل (2) يمثل معمارية شبكة الانحدار العام العصبية GRNN



شكل (2) يمثل معمارية شبكة الانحدار العام العصبية GRNN

منهجية شبكة الانحدار العام العصبية GRNN:

تتضمن هذا الشبكة في تركيبها على مرحلتين رئيسيتين من مراحل المعالجة للمعلومات هي

كالاتي: [9-11]

المرحلة الاولى:

في هذه المرحلة يتم ايجاد المسافة بين قيم المدخلات والتي غالباً ما تكون بشكل متجه وقيم الاوزان (والتي غالباً ما تكون متجه ايضاً)، ثم يضرب الناتج بقيمة التحيز (Bias) وهو

مقدار صغير. وتكون دالة التنشيط من نوع كاوسين وهي من الدوال غير الخطية التي تأخذ الصيغة الآتية:

$$f(x) = e^{-x^2} \dots \dots \dots (6)$$

المرحلة الثانية:

تسمى بالطبقة الخطية الخاصة (Special Linear) ويتم فيها إيجاد الضرب النقطي بين متجه الأوزان ومتجه الإدخال (الذي يمثل إخراج المرحلة الأولى "السابقة")، ثم يقسم الناتج على مجموع قيم الإدخال ولا يتم فيها حساب قيمة التحيز، أو حد العتبة ودالة التحويل لهذه الطبقة هي دالة خطية تأخذ الصيغة الآتية:

$$f(x) = x \dots \dots \dots (7)$$

الجانب التطبيقي:

توصيف البيانات:

تضم عينة البحث مجموعتين من البيانات حيث اعتمدت كمية إنتاج المحاصيل الزراعية الاستراتيجية كمتغيرات للمجموعة الأولى Y's (y₁, y₂, y₃) وتمثل كمية إنتاج الحنطة، الشعير، الرز على التوالي للفترة من 1978-2006 أما متغيرات المجموعة الثانية X's (x₁, x₂, x₃, x₄) وتمثل "المتوسط العام للمساحة المزروعة (حنطة، شعير، رز) المعدل العام لسعر (الحنطة، الشعير، الرز)، كمية المياه السطحية، معدل النمو السنوي للسكان" على التوالي للفترة من 1978-2006. والجدول (1) يبين بيانات البحث. علماً بأن البيانات جمعت من السجلات الخاصة بالجهاز المركزي للإحصاء واحتسبت الأسعار بالدولار لتلافي حالة التضخم في الأسعار بعد الحصار.

جدول (1) يمثل بيانات البحث

	Y1	Y2	Y3	X1	X2	X3	X4
1978	9998	6172	1720	32327	200.5333	80543	3
1979	6848	5714	1578	28909.33	228.2667	69820.7	3
1980	9056	6824	1465	31468.33	249.6	84958	3
1981	9020	9247	1622	30613.67	276.2667	81678.3	3
1982	9651	9020	1634	31853.67	309.3333	84548	3
1983	8410	8355	1105	36188.67	353.0667	70798.3	3
1984	4705	4820	1087	37320.33	354.1333	50110.7	3
1985	13023	13314	1589	41020.33	426.6667	90602.9	3
1986	9991	10459	1412	37865.67	458.6667	44718	3
1987	7222	7429	1959	36658.33	608	74016	2.9
1988	9292	14368	1506	34928	608	152350	2.9
1989	6022	6634	2318	33470.67	63.33333	46452.5	2.9
1990	11958	18543	2288	43675.33	81.2	45538	2.9
1991	14764	7684	1891	66907	76.66667	43270	2.9
1992	13107	13538	2573	41782.33	158.7333	74270	2.9

1993	9110	8904	2719	38651.67	65.33333	78730	2.9
1994	9821	8540	3829	37606.67	51.03333	60180	2.9
1995	10914	7124	3123	36767	112.5	89530	2.9
1996	11500	6473	2829	33898.67	102.7333	68850	2.9
1997	10421	4296	2743	33195.33	87.26667	70300	3
1998	11102	6299	3895	35221	59.4	37410	3
1999	11216	2830	2185	34941	67.6	36080	3
2000	10403	1927	124	22083	79.43333	36080	3
2001	13657	7128	49	24809	89	30520	3
2002	19482	8334	1938	35578.33	86	52910	3
2003	19832	8604	813	37434.33	120.9667	73090	3
2004	18321	8054	2503	34465.33	206.4667	64460	3
2005	18884	7544	3087	36970	288.8667	55300	3
2006	17532	9193	3633	35530.67	341.8667	63100	3

المصدر : الجهاز المركزي للإحصاء

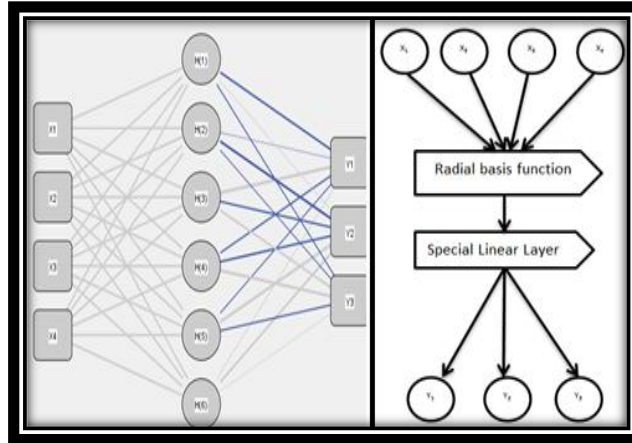
المعالجة:

لبيان اثر متغيرات المجموعة المستقلة في متغيرات المجموعة المعتمدة لنموذج متعدد المتغيرات باستعمال منهجية شبكة الانحدار العام GRNN، ان معمارية هذه الشبكة لنموذج متعدد المتغيرات تتكون من ثلاثة طبقات رئيسية ويكون الاختلاف في شبكة بعدد العقد في طبقة المدخلات وحسب نوع التجربة لقياس اي المتغيرات المستقلة الاكثر تأثيراً وكانت كالآتي:

طبقة المدخلات: التجربة الاولى تمثل اربعة عقد وهي متغيرات المجموعة الثانية من عينة البحث وهي (X_1, X_2, X_3, X_4) التجربة الثاني تمثل عقدة واحدة هو المتغير X_1 ، التجربة الثالث تمثل عقدة واحدة هو المتغير X_2 ، التجربة الرابعة تمثل عقدة واحدة هو المتغير X_3 ، التجربة الخامسة تمثل عقدة واحدة هو المتغير X_4

الطبقة الوسطى: تمثل عمليات دالة الاساس الاشعاعي.

طبقة المخرجات: تمثل ثلاثة عقد وهي متغيرات المجموعة الاولى من عينة البحث وهي (y_1, y_2, y_3) . والشكل الاتي يبين معمارية شبكة الانحدار العام لبيانات البحث:



شكل (3) يمثل معمارية شبكة الانحدار العام

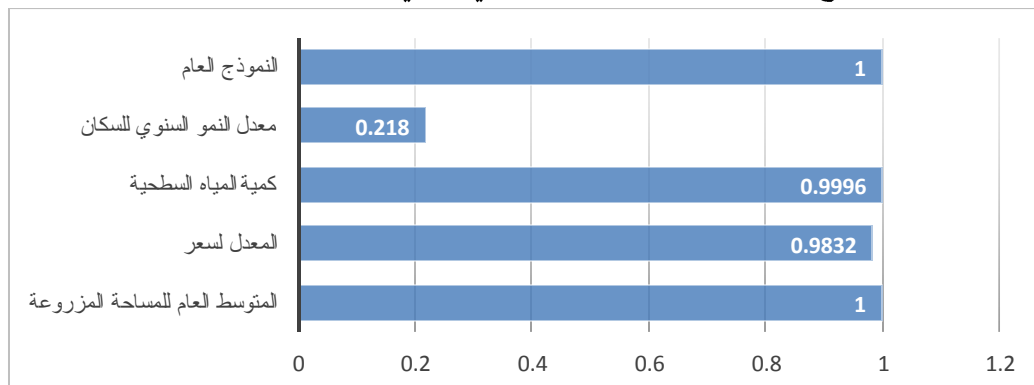
النتائج:

يمكن تلخيص نتائج اربعة تجارب لنموذج متعدد المتغيرات باستعمال شبكة الانحدار العام استناداً الى قياس معامل اللافاضلة لغرض تشخيص اي المتغيرات اكثر تأثيراً التي استخرجت بواسطة برنامج ماتلاب باستخدام الدالة newgrnn لإنشاء الشبكة ودالة Sim لمحاكاة الشبكة وحساب المخرجات في الجدول رقم (2).

جدول (2) يمثل نتائج تقييم الشبكة العصبية الاصطناعية

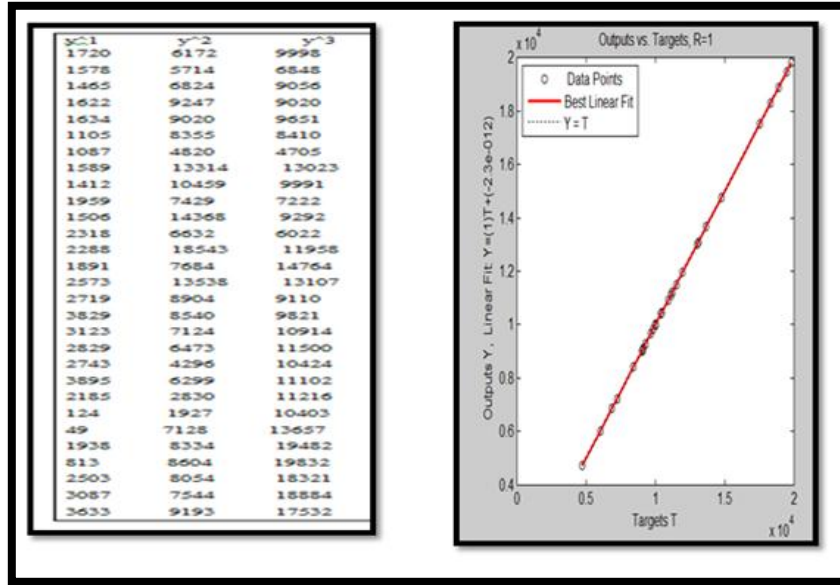
المتغيرات المؤثرة المعيار	X_1 المتوسط العام للمساحة المزروعة	X_2 المعدل لسعر	X_3 كمية المياه السطحية	X_4 معدل النمو السنوي للسكان	X_1, X_2, X_3, X_4 النموذج العام
Rd	1	0.9832	0.9996	0.218	1

ويمكن تمثيل النتائج بياناً من خلال الشكل البياني الآتي:



شكل (4) يمثل المدرج التكراري للنتائج النهائية

والشكل الاتي يبين تقييم الشبكة ومخرجاتها لنموذج قيد البحث. والشكل الاتي يبين نتائج تقييم الشبكة العصبية لنموذج العام.



شكل (5) يمثل نتائج تقييم ومخرجات الشبكة

تحليل النتائج:

يتبين من نتائج النهائية المبينة في جدول (2) ومن خلال نتائج شبكة الانحدار العام، ان المتغيرات التي تأثير معنوي على نموج متعدد المتغيرات قيد البحث التي تحقق اعلى قيمة معامل افاضة و اقل خطأ و افضل توفيق ولديه تأثير تام هي المتوسط العام للمساحة المزروعة، كمية المياه السطحية على توالي حيث كانت قيمة معامل الافاضلة 1 تقريبا اما المتغير الاقل تأثير هو معدل النمو وكانت قيمة معامل الافاضلة 0.2 وهي قيمة متدنية ويتبين ايضا ان كفاءة مجموعة المتغيرات المستقلة الاربعة على النموذج حيث كانت قيمة معامل الافاضلة 100% تامة اي تمكنت المتغيرات التوضيحية "المستقلة" من تفسير النموذج بشكل تام اي 100%.

الاستنتاجات:

يمكن ايجاز اهم النتائج والاستنتاجات التي التوصل اليها من خلال هذا البحث بالاتي:

1. اثبت النتائج النهائي كفاءة الشبكات العصبية الاحصائية بشكل خاص والشبكات العصبية الاصطناعية بشكل عام الشبكات العصبية الاحصائية بشكل تام او كبير في معالجة نماذج متعدد المتغيرات ودراسة تأثير المتغيرات المستقلة.

2. اثبتت النتائج النهائية بان المتغيرات الاكثر تأثير على كمية انتاج الحنطة، الشعير والرز هي المتوسط العام للمساحة المزروعة بشكل تام والمعدل العام لسعر، كمية المياه السطحية بشكل كبيرة جدا اما معدل النمو السكان له تأثير قليلة على النموذج او المحاصيل الزراعية الاستراتيجية لاتخاذ قرارات سلمية ورصينة وبناء خطط استراتيجية كفوءة وذات جودة عالية.

3. اعتماد الشبكات العصبية الاحصائية بشكل خاص والشبكات العصبية الاصطناعية بشكل عام كبديل عن الطرائق التقليدية المعتمدة لمعالجة نماذج متعدد المتغيرات كالتحليل القويم وخصوصا في العلاقات او النماذج او شبه خطية او غير خطية او حتى الخطية، مما يوفر اداة قوية ومرنة وتتصف بالذكاء لمعالجة نماذج متعدد المتغيرات بكفاءة تامة وجودة عالية في حين النماذج التقليدية تتطلب بناء نموذج رياضي الذي يحتاج مهارات رياضية وخبرة كبيرة.

المصادر والمراجع:

1. عبد الله، سهيلة نجم.دراسة تحليل الارتباط القويم لدراسة تأثير مجموعة من العوامل على انتاج المحاصيل الاستراتيجية. مجلة الادارة والاقتصاد-الجامعة المستنصرية.2008. عدد 73.ص198-214.

2. Alexander, I and Galushkin, G. *Neural Networks Theory*. Springer-Verlag London UK PP884.2012

3. Chae, Y.T.; Horesh, R.; Hwang, Y.; Lee, Y.M. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy Build.* 2016, 111, 184–194.

4. Dully, M. *Traffic Safety Evaluation of Future Road Lighting Systems*. Master's Thesis, Linköping University, Linköping, Sweden, 2013.

5. Ji, Y.; Xu, P.; Duan, P.; Lu, X. Estimating hourly cooling load in commercial buildings using a thermal network model and electricity submetering data. *Appl. Energy* 2016, 169, 309–323.

6. Kim,W.; Jeon, Y.; Kim, Y. Simulation-based optimization of an integrated daylighting and HVAC system using the design of experiments method. *Appl. Energy* 2016, 162, 666–674.

7. Makarynsky, O. Improving wave predictions with artificial neural networks. *Ocean Eng.* 2004, 31, 709–724.

8. Moon, J.W.; Chung, S.K. Development of a thermal control algorithm using artificial neural network models for improved thermal comfort and energy efficiency in accommodation buildings. *Appl. Therm. Eng.* 2016,103, 1135–1144.

9. Papantoniou, S.; Kolokotsa, D. Prediction of outdoor air temperature using neural networks: Application in 4 European cities. Energy Build. 2016, 114, 72–79.
10. Rencher, Alvin C. Methods of Multivariate Analysis. Second Edition. John Wiley & Sons, Inc.2012
11. Sahin, M.; Oğuz, Y.; Büyüktümtürk, F. ANN-based estimation of time-dependent energy loss in lighting systems. Energy Build. 2016, 116, 455–467.