



## مقارنة بين الخوارزمية الجينية MWCD والشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية) لتحديد المؤشرات المسببة لتلوث المياه

فاطمة عبد الحميد جواد<sup>a\*</sup> ، صباح منفي<sup>b</sup>  
جامعة بغداد /كلية الادارة والاقتصاد/ قسم الاحصاء

### المخلص

### معلومات المقالة

تستعمل الطرق اللامعلمية في البيانات التي تحتوي على قيم شاذة، الاهمية الاساسية في استعمال الطرق اللامعلمية هو تحديد موقع الوسيط، ففي انموذج الانحدار متعدد المتغيرات يكون من الصعوبة تحديد موقع الوسيط لوجود اكثر من بعد وتشتت القيم وزيادة بيانات الظاهرة المدروسة. فقد تم تطبيق الخوارزمية الجينية Minimum Weighted Covariance Determinant Estimator وتكتب باختصار (MWCD) ومقارنتها مع الشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية) (Back Propagation Algorithm) Multilayer Network في ايجاد التقدير لموقع الوسيط بالاعتماد على مسافة مهلنوبس Mahalanobis Distance (MD) واصغر محدد لمصفوفة التباين المشترك Minimum Covariance Determinant (MCD) كواحدة من الطرق اللامعلمية الحصينة. وقد طبقت الدراسة على الاحصاءات البيئية لتلوث مياه الشرب لسنة 2013 شملت محافظات العراق كافة عدا اقليم كردستان، مقسمة الى 10 اشهر وقد استخدمت 9 انواع من المؤشرات الكيميائية والفيزيائية المسببة لتلوث مياه الشرب عند تجاوز الحد الاعلى للقياس. ولتحديد المؤشرات المسببة للتلوث تم استخدام شرائح التمهيد Smoothing Spline في تقدير معاملات انموذج الانحدار متعدد المتغيرات للمعالم المتغيرة زمنيا وقد قدرت المعالم الممهدة بطريقة تثبيت التقاطع Cross Validation (CV). اثبتت نتائج المقارنة بفاعلية الشبكة العصبية متعددة الطبقات الارجاعية.

### تاريخ البحث

الاستلام :

تاريخ التعديل :

قبول النشر : 2020/11/19

متوفر على الأنترنت: 2021/8/29

### الكلمات المفتاحية :

الشبكة العصبية

مهلنوبس

شرائح التمهيد

الخوارزمية الجينية

الشبكات متعددة الطبقات



A comparison between the genetic algorithm MWCD and the Multi-Layered (Back Propagation Algorithm) network to identify the indicators causing water pollution

Fatimah Abdul Hmeed Jawad<sup>a\*</sup> ، Sabah manfi redha<sup>b</sup>  
University of Baghdad/college of Economics & Administration/department of statistics.

### Abstract

Nonparametric methods are used in the data that contain outliers values. The main importance in using Nonparametric methods is to locate the median in the multivariate regression model . It is difficult to locate the median due to the presence of more than one dimension and the dispersion of values and the increase of the studied phenomenon data .The genetic algorithms Minimum Weighted Covariance Determinant Estimator (MWCD), was applied and compared with the multilayer neural network Back propagation to find the estimate of the median location based on the minimum distance (Mahalanobis Distance) and smallest specified for the variance matrix . Joint Minimum Covariance Determinant (MCD) as one of the most nonparametric methods robust .

The study has been applied on environmental pollution statistics of drinking water for the year (2013) including all the Iraqi provinces except Kurdistan region , divided into (10) months .To determine the contaminating indicators ,Smoothing Spline slides were used to estimate the multivariate regression model parameters for the time – variable parameters . The parameters smoothed by the cross validation (CV ) method were estimated the results of the comparison proved the effectiveness of the retrospective Multi – Layer neural network.

**Key words:** genetic algorithm , neural network , mahalanobis ,smoothing spline , multiLayer networks.

\*

Corresponding author : E-mail addresses .:

2021 AL – Muthanna University . DOI:10.52113/6/2021-11/290 -304.

## المقدمة

في عام ( 2009 ) قام الباحث Stefan و Ella Roelant و Van Aelst و Gert Willems بتقدير الموقع والتشتت لبيانات متعدد المتغيرات مع مجموعة من الاوزان بالاعتماد على رتبة المسافة Mahalanobis ، حيث تم استخدام دالة الوزن  $a_n$  وقد فسر ان مقدر MWCD اكثر عموما من مقدر (MCD) والاعتماد على دالة الوزن يمكن ان يعالج المتوسط لمختلف القيم المتطرفة.

### الإطار النظري

#### أولاً: مفهوم الإدارة المفتوحة

- التقديرات اللامعلمية الممهدة لانموذج المعلمات المتغيرة زمنياً للبيانات الطولية

#### - Nonparametric Smoothing Estimates of time Varying Coefficient Models with Longitudinal data

سيتم في هذا المبحث تقدير معلمات الانموذج المتغيرة زمنياً (Z.Huang, Colin O.wu, & Lan Zhou, 2004) (بيانات القياسات المتكررة – repeated measurement) الطولية (Longitudinal) التي توضح تأثير متغيرات الاستجابة خلال الزمن (t) حسب الاشهر او السنوات ، فقد تم استعمال التقديرات اللامعلمية لتحليل البيانات ذات القياسات المتكررة (الطولية) وذلك باستعمال شرائح التمهيد (Smoothing Spline) لتقدير المعلمات  $\beta(t)$  (R.Hoover, John A.Rice, Colin O.wu, & LI-PING YANG, 1998) ويمكن توضيح أنموذج المعلمات المتغيرة زمنياً (Time Varying Coefficient) وتكتب بأختصار (TVC) كالآتي :-

$$Y(t) = X'(t) B(t) + \varepsilon(t) \dots (2-1)$$

اذ ان :-

$n_i$  تمثل  $i=1, \dots, n$  عدد القطاعات ، اما  $j=1, \dots, n_i$  تمثل  $n_i$  عدد المشاهدات للقطاع  $i^{th}$  .

$t_{ij}$  : الزمن للقياسات  $j^{th}$  للقطاعات  $i^{th}$  .

$Y(t)$  : المتغير المعتمد  $(1 * n_i)$  متجة الاستجابة  $j^{th}$  .

$X(t)$  : مصفوفة المتغيرات المستقلة (التوضيحية)  $(n_i * d)$  للقطاع  $i^{th}$  .

$d$  : تمثل عدد المتغيرات المستقلة  $(d=1, \dots, r)$  .

لتقدير نموذج الانحدار متعدد المتغيرات في حالة زيادة عدد المشاهدات  $n$  وعندما تكون عدد المتغيرات  $p \geq 2$  وتشتت البيانات ووجود اكثر من بعد وكذلك وجود القيم الشاذة (المتطرفة) اذ ان الطرق التقليدية تتأثر في التقدير ، فقد تم استعمال الطرق اللامعلمية في التقدير والتي تعتمد على تحديد موقع الوسيط بالاعتماد على مسافة Mahalanobis Distance(MD) والتي يتم ايجادها من المتغيرات التوضيحية وعلى اصغر محدد لمصفوفة التباين المشترك Minimum Covariance Determinant(MCD) ، وقد تم تطبيق الخوارزمية الجينية MWCD والشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية) Back Multilayer Network (Propagation Algorithm) للمقارنة من حيث الدقة في النتائج والسرعة في ايجاد موقع الوسيط ، ثم تقدير معالم النموذج لتحديد المؤشرات المسببة لتلوث مياه الشرب .

من الباحثين الذين استخدموا الخوارزمية الجينية MWCD في عام (2003) م قام الباحثون (Peter J.Rousseeuw و Van Aelst Stefan و Katrien Van Driessen و Jose Agullo) بتقديم الطرق الحصينة للتقدير لانموذج انحدار متعدد المتغيرات مستندا بذلك على الموقع ومصفوفة الانتشار باستعمال المتغيرات التوضيحية والاستجابة في التقدير ، وقد استعمل اسلوب المحاكاة لمختلف احجام العينات ولزيادة الكفاءة اقترح عدة اوزان للتقدير ، وهي اوزان الوسيط الموقوع (L-reweighting location) واوزان للقيم الاولية لانموذج الانحدار (reweighted regression estimators – R) وكانت اكثر كفاءة بالمقارنة بينها وبين التقديرات الى (MCD) ، اما عند الدمج في حالة الـوزن للموقع والانموذج (LR-weighted MCD) اضافت خصائص للتحيز bias متوسط مربع الخطأ (MSE) اقل قيمة بالمقارنة مع التقدير بطريقة (Least square regression) .

في عام (2003) م قدم الباحثان (Greet Pison و Stefan Van Aelst) اقترحت عدة مقدرات حصينة منها محدد اصغر لمصفوفة التباين المشترك (MCD) الخطوة الاولى في تقدير موقع الوسيط هو اعادة الوزن الى الوسيط الحسابي ومصفوفة التباين المشترك وذلك لزيادة الكفاءة وسميت اعادة وزن محدد اصغر لمصفوفة التباين المشترك (Reweighted Minimum Covariance Determinant (RMCD) ، وتم اختيار نقطة التوقف 50% عندما تكون  $\gamma = 0.5$  اما نقطة التوقف 25% عندما تكون  $\gamma = 0.75$  .

$\beta(t)$  : دوال المعالم التمهيدية في الزمن  $(t_{ij})$  متجة (  $n_i * 1$  )  
 مقدر  $B(t)$  باستعمال جميع المشاهدات ما عدا تلك المشاهدات من القطاع  $i$  .

$\varepsilon(t)$ : متجة الخطأ العشوائي  $(n_i * 1)$  .  
 يمكن اعادة كتابتها كالآتي :-

$$\dots(2-2) y_{ij} = x'_{ij} \beta(t_{ij}) + \varepsilon_i(t_{ij})$$

$(n_i * 1)$  متجة  $y_i = (y_{i1}, \dots, y_{in_i})'$  ,  $(n_i * 1)$  متجة  
 $Y = (y'_1, \dots, y'_n)'$

$X = (x'_1, \dots, x'_n)$  مصفوفة  $(n_i * dn_i)$  مصفوفة  $(n_i * dn_i)$  ,  
 $x_i = (x_{i1}, \dots, x_{in_i})'$

### - خصائص التقدير بطريقة الشرائح الممهدة (Smoothing Spline)

ان تكون دوال المعالم  $(\beta_1(t), \dots, \beta_k(t))$  مستمرة وقابلة للاشتقاق ، المشتقة الثانية لها

$(\beta''_1(t), \dots, \beta''_k(t))$  محددة ويمكن تكامل مربع المشتقة (Brabanter, Johan A.K.suykens, & Bart De Moor, 2013) .

اما تقدير المعالم يكون بتصغير دالة المربعات الصغرى الجزائية  $J(\beta, \lambda)$  :-

$$J(\beta, \lambda) = \sum_{i=1}^n \sum_{j=1}^{n_i} [y_{ij} - \{\sum_{r=1}^d x_{ijr} \beta_r(t_{ij})\}]^2 + \sum_{r=1}^d \lambda_r \int \{\beta''_r(t)\}^2 dt \dots(2-3)$$

وان :-

$\lambda_r = (\lambda_1, \dots, \lambda_d)'$  تمثل قيم موجبة للمعلم الممهدة تتغير بتغير دوال المعاملات ، ومن اهم شرائح التمهيد هي شرائح التمهيد التكعيبيية (cubic smoothing splines) (css)

يتم ايجاد التكامل الذي يعرف بالجزء غير الممهدة وتقدير المعالم الذي يتزايد عددها بزيادة عدد المشاهدات.

ولاختيار معلمة التمهيد (smoothing parameter selection) سيتم اختيار معيار العبور الشرعي (Cross Validation) وتكتب باختصار (CV) كالآتي :-

$$CV(\lambda) = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{n_i} \{y_{ij} - x'_{ij} \hat{\beta}^{(-i)}(t_{ij})\}^2 \dots(2-4)$$

وان :-

لذا سيتم استخدام طريقة تقدير المرحلتين (Two – Step Estimation Method) لموائمة أنموذج المعاملات المتغيرة زمنياً (P.P.B.Eggermont, R.L.Eubank, & V.N.Lariccia, 2005) ، فأن طريقة css تستعمل  $d$  من معالم التمهيد يتم استعمال خوارزمية Backfitting على الرغم انها تحل مشكلة تعدد الابعاد لكنها تعاني من مشكلة اذ ان حجم المصفوفة يزداد بزيادة عدد القطاعات  $n$  او بزيادة عدد نقاط الزمن المحددة  $n_i$  والذي يؤدي الى زيادة معالم التمهيد ، فأن النهج البديل هو اجراء تقدير المرحلتين .

يتم اولاً حساب المقدرات الخام لموائمة أنموذج خطي قياسي ، وثانياً نمهد المقدرات الخام على المقدرات التمهيدية لدوال المعاملات .

إن اجراء المرحلتين (Two – step) (WANG, Hongzhe & Jianhua UANG, 2008) يمتلك خصائص جيدة كثيرة، فهي بسيطة الفهم، سهلة التنفيذ، سريعة الحساب وفعالة في الاداء، إن هذا الاجراء له دوافع بواسطة هيكل خاص من العديد من مجموعات البيانات الطولية، حيث يتم جمع القياسات على نفس نقاط الزمن المحددة الى جميع القطاعات (على الرغم من إن عدد القياسات لكل قطاع غير متساوية او حتى مفقودة).

لنفترض  $t_j, j=1, \dots, n_i$  هي نقاط زمن محددة، حيث تم جمع البيانات. لأن هناك عدد من المشاهدات التي جمعت في الزمن

$t_j$  ، فمن الممكن لهذا الثابت  $t_j$  استعمال البيانات المجمعة هناك لموائمة أنموذج المعاملات المتغيرة زمنياً والحصول على المقدرات الخام (Raw estimates) .

$$b(t_j) = (b_1(t_j), \dots, b_d(t_j))'$$

$$B(t_j) = (B_1(t_j), \dots, B_d(t_j))'$$

هذه هي المرحلة الأولى، عادةً المقدرات الخام هي غير ممهدة تحتاج الى تمهيدها للحصول على المقدرات الممهدة الى دوال المعاملات لذلك، في المرحلة الثانية لكل مركبة معطاة  $r = 1, 2, \dots, d$  نطبق تقنية تمهيد الى البيانات  $\{b_r(t_j), t_j\}, j = 1, 2, \dots, m$

n : تمثل عدد المشاهدات

h : تمثل اعلى نقطة توقف  $h = \frac{n+p+1}{2}$

اي ان مقدر (MWCD) اكثر عموماً من مقدر (MCD) ويتم فيها خفض دالة الوزن ، يضع وزنا اكبر على النقاط القريبة من المركز ، في حين ان وزن الدالة يعطي وزنا اكبر لمزيد من النقاط البعيدة عن المركز ، الاعتماد على وظيفة الوزن (MWCD) تعالج المتوسط لمختلف القيم المتطرفة (Zuo, 2004) .

لتقدير المركز M يتم ايجاد دالة الوزن كالآتي :-

$$a_n(i) = h^+ \left( \frac{n_i}{(n+1)} \right) , \dots (2-5)$$

$$i = 1, \dots, n$$

$$h^+(0, 1) \Rightarrow [0, \infty)$$

$$\text{Sup} \{u; h^+(u) > 0\} = 1 - \alpha , 0 \leq \alpha \leq \frac{1}{2}$$

$$\text{Sup} : \text{اعلى قيمة الى } h^+ \text{ عندما تكون } 0 \leq \alpha \leq \frac{1}{2}$$

$$h^+(u) > 0 \text{ for } u \in (0, 1-\alpha]$$

$\alpha$  النسبة للمشاهدات  $x_i$  تأخذ الوزن صفر وبذلك نحصل على التقدير الحصين .

اما الخوارزمية المستعملة لايجاد (MWCD) التقريبية والمشابه الى خوارزمية (MCD) وذلك باعتبار دالة الوزن غير متزايدة وأجراء بعض التعديلات وتعتبر تعميم الى (c – step) كالآتي :-

$$X_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^p$$

دالة الوزن  $a_n$  دالة غير متزايدة موضحة في المعادلة (2-5) .

الخطوة الاولى :

$$Q_1 = \sum_{i=1}^n a_n R_{1i} MD_1^2(i) \dots (2-6)$$

$a_n$  : تمثل دالة الوزن

$MD_1^2(i)$  : تمثل مسافة مهلنوبس

$R_{1i}$  : تمثل الرتبة الى مسافة مهلنوبس

- نجد المسافة

$$MD_1^2(i) = (x_i - \hat{M}_1)' V_1^{-1} (x_i - \hat{M}_1)$$

ان مرحلة تقديرات التمهيد (Smoothing estimates) هذه حاسمة لأنها تعطي مقدرات تمهيدية لدوال معاملات التمهيد الاساسية، وإضافة الى ذلك فان مرحلة التمهيد عادة ذات بعد واحد (one-dimensional) .

ان إجراء المرحلتين سهل التنفيذ ، حيث لكل ثابت  $t, j$  ، أنموذج (TVC) هو أنموذج خطي قياسي مع هيكل أخطاء مستقل، وهذا الاجراء سريع في الحساب، وان الاسباب الرئيسية لسرعة الحساب، في المرحلة الاولى الحساب يركز فقط على نقطة معينة وبالتالي البيانات المعنية هي قليلة جداً بالمقارنة مع بيانات المجموعة الكاملة، وفي المرحلة الثانية يتم تنفيذ العملية الحسابية لمجرد تمهيد عدة مشاكل ذات بعد واحد .

#### - مقدمة في الخوارزمية الجينية

بين كل من (شعبان، حسن ياسين طعمة، و حسن ثابت كرماشة، 2009) الخوارزمية هي مجموعة محددة من خطوات الحل التي تؤدي الى انجاز وظيفة معينة ويجب ان تتوفر فيها الشروط التالية:-

1- المدخلات (INPUT) : صفر او اكثر من قيمة .

2- المخرجات (OUTPUT) : قيمة واحدة على الاقل .

3- الوضوح (DEFINITENESS) : كل خطوة في الخوارزمية واضحة المعاني وغير غامضة اي يجب ان تفهم من قبل الجميع .

4- المحدودية (FINITENES) : كل خطوات الخوارزميات يمكن حلها في فترة زمنية محددة .

5- المحلولة (EFFECTIVENESS) : كل خطوة تكون ممكنة الحل .

#### - خوارزمية أصغر محدد للتباين المشترك الموزون (Minimum Weighted Covariance Determinant)

تستخدم المقدرات المرجحة لتقدير الموقع لتشتت بيانات متعدد المتغيرات بالاستناد الى رتبة مسافة (Mahalanobis Distance) (Roelant, Stefan Van Aelst, & Gert, 2009) حيث قام باعطاء تقديرات المتوسط ومصفوفة التباين المشترك الى نصف البيانات يتم فيها الحصول على اصغر محدد لمصفوفة التباين المشترك (MCD) والتي تكون فيها دالة الوزن مساوية الى الصفر ، لذا فإن دالة الوزن اكثر عمومية واكثر ملائمة يمكن توضيح دالة الوزن كالآتي :-

$$a_n(i) = I , (i \leq h) , \frac{n}{2} \leq h \leq n$$

$$i=1, \dots, n \quad \dots(2-7)$$

$$Q_2 = \sum_{i=1}^n a_n (R_{2i})^2 d_2^2(i) \quad \dots(2-13)$$

$$Q_2 \leq Q_1 \text{ اذا كان } \hat{M}_2 = \hat{M}_1, \hat{V}_2 = \hat{V}_1$$

$a_n$  : تمثل دالة الوزن

$$MD_2^2(i) : \text{ تمثل مسافة مهلوبس}$$

$R_{2i}$  : تمثل الرتبة الى مسافة مهلوبس

بالاستناد الى الخوارزمية العامة c-step الموضحة كالآتي :-

1- نبدأ بسحب عشوائي (n+1) , مجموعه جزئية  $j_m$  الى  $x_n$  .

2- نحسب متوسط العينة المطابقة  $\hat{M}_m$  ومصفوفة التباين المشترك  $\hat{V}_m$

اذا  $|\hat{V}_m| = 0$  لكل  $j_m$  نضيف نقطة الى  $j_m$  الى ان نحصل على  $|\hat{V}_m| > 0$  او  $(j_m = n)$  .

3- لكل مجموعة جزئية نحسب الدالة  $Q_1$  بالاعتماد على  $(\hat{M}_m, \hat{V}_m)$  .

4- نطبق الخطوة 2 من c-step .

5- نختار 10 مجموعة جزئية والتي تعطي اقل قيمة الى الدالة  $Q_1$  ننفذ مرة اخرى (c-step) .

6- الحل النهائي يقرر بواسطة الخوارزمية الى  $\hat{M}$  و  $\hat{V}$  مطابقة الى اقل قيمة الى الدالة  $Q_1$  بين القيم (10) .

تحدد العينة الوسيطة تكون لديها اقل قيمة الى الدالة  $Q_1$  .

ملاحظة بسبب اختيار عدد من التوافيق الى رتبة المتجه R لذا يمكن تحديد عدد من المعدل الموزون والتباين المشترك الموزون بالمعادلة (2-8) و (2-9) .

$\hat{M}_1 \in \mathbb{R}^{p \times p}$  تمثل الوسط الحسابي و  $\hat{V}_1 \in \mathbb{R}^{p \times p}$  تمثل مصفوفة التباين المشترك و  $|\hat{V}_1| = 1$

- نجد الرتبة  $R=(R_{11}, \dots, R_{1n})$  الى متجه المسافة  $MD_1^2(i)$

الخطوة الثانية :

- نجد الوسط الحسابي الموزون ومصفوفة التباين المشترك الموزون كالآتي :-

$$\hat{M}_2 = \hat{M}(R_1) = \frac{\sum_{i=1}^n a_n(R_{1i})x_i}{\sum_{i=1}^n a_n(R_{1i})} \quad \dots(2-8)$$

$$\hat{\Sigma}_2 = \hat{\Sigma}(R_1) = c_h^+ \frac{\sum_{i=1}^n a_n(R_{1i})(x_i - \hat{M}(R_{1i}))(x_i - \hat{M}(R_{1i}))'}{\sum_{i=1}^n a_n(R_{1i})} \quad \dots (2-9)$$

وان :-

$c_h^+$  عامل متسق (Chakraborty (Fisher – consistency) Brobal Chaudhuri, 2005) حيث ان :-

$$c_h^+ = \alpha / F(\chi_{p+2}^2), (\chi_p^2) \quad \dots(2-10)$$

الخطوة الثالثة :

$$\hat{V}_2 = |\hat{\Sigma}_2|^{-\frac{1}{p}} \hat{\Sigma}_2 \quad \dots (2-11) \quad \text{نجد}$$

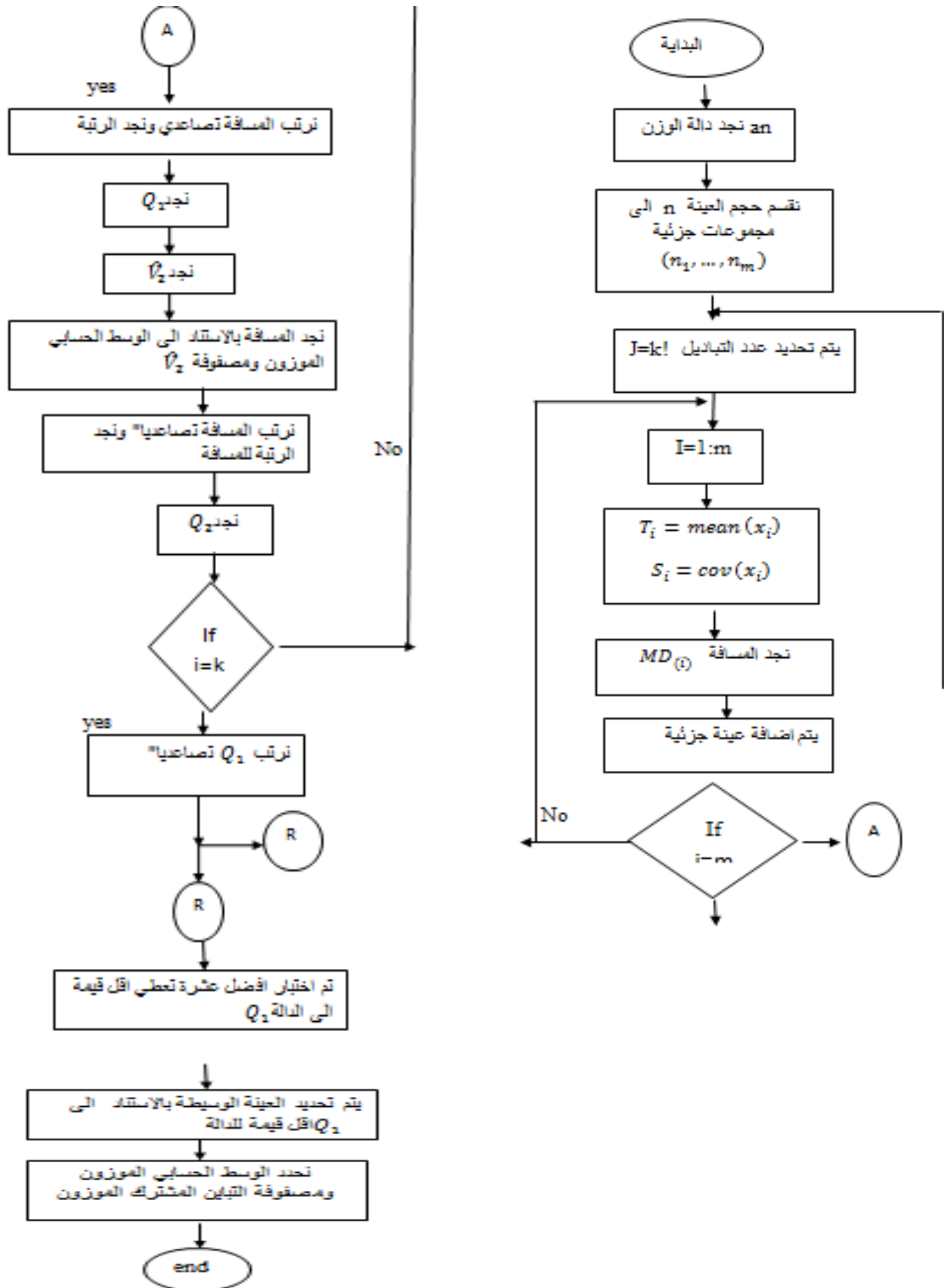
- نجد مسافة مهلوبس بالاستناد الى الوسط الحسابي  $\hat{M}_2$  ومصفوفة التباين المشترك  $\hat{V}_2$

$$MD_2^2(i) = (x_i - \hat{M}_2)' \hat{V}_2^{-1} (x_i - \hat{M}_2) \quad \dots(2-12)$$

$i=1, \dots, n$

- نجد الرتبة  $R_2$  الى المسافة  $MD_2^2(i)$

الشكل (1) المخطط الانسيابي لخوارزمية MWCD



## - مقدمة في الشبكات العصبية

- المحاكاة هي العملية الاساسية للعصبون وهناك عمليات اخرى اساسية لعملية المعالجة في الشبكة العصبية هي :-

- 1- neurons / node
- 2- connection /link
- 3- weights
- 4- summation
- 5- activation function

اهم نقطة في الشبكات العصبية هي الاوزان وهي الحل الامثل لحل المشكلة بالاضافة الى معمارية الشبكة Activation Function الميزة الاساسية للتفريق بين انواع الشبكات العصبية ، شبكة عصبية وحيدة الطبقة Linear وشبكة عصبية متعددة الطبقات Non Linear .

## - الشبكات متعددة الطبقات MultiLayer Networks

تمتلك هذه الشبكات طبقة واحدة او اكثر من العقد المخفية Hidden Nodes

هذه العقد لا تمثل وحدات ادخال او اخراج ، ولها استخدامات عديدة مثل تمييز النماذج ، تمييز الاصوات ، تحليل البصمة الصوتية ، في خوارزميات الضغط والخوارزميات الجينية... الخ.

كدالة تحويلية ( sigmoid function ) وتستخدم الدالة الاسية وهناك ثلاث (Despagne & D.Luc massart, 1998) انواع من الدوال :-

$$1- F(x) = \tanh(x/B)$$

$$2- F(x) = \frac{e^{Bx} - e^{-Bx}}{e^{Bx} + e^{-Bx}}$$

$$B > 0 \quad 3- F(x) = \frac{1}{1 + e^{-Bx}}$$

وقبل البدء باستخدام الشبكات العصبية لا بد من تحديد الاتي :-

1- العتبة Threshold

2- الدالة التفعيلية Activation function

الشبكات العصبية يمكن ان تعرف بأنها نظم او تقنية لمعالجة المعلومات منبثقة من دراسة المخ والنظام العصبي للانسان ، او هو نظام معالجة معلومات يحاكي اسلوب الشبكات العصبية الحيوية وله مميزات اداء معينة او خوارزميات او نماذج رياضية لمعالجة المعلومات .

تتكون الشبكة العصبية من مجموعة من العصبونات او الوحدات او الخلايا او العقد موزعة في طبقات وهي (D.crawford & Rogerl.wainwright, 1995) :-

- 1- طبقة الدخل
- 2- طبقة الخرج
- 3- طبقة المعالجة

ترتبط العقد مع بعضها البعض بروابط اتصال وترفق بها قيم عددية ( اوزان ) وتعتبر المعلومات الاساسية لحل المشكلة وفيها يبدأ الحل حيث تستقبل طبقة الدخل الاشارة وتعمل على توزيعها الى الطبقات الاخرى حسب نوع الاشارة (Wu,j.1994) .  
- الخلية العصبية تستقبل الاشارة وتعالجها ونحصل على مخرجات

Input → process → output

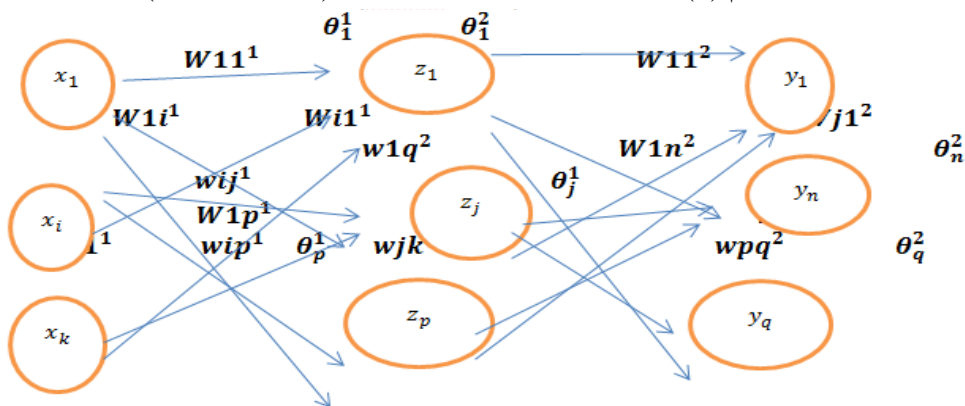
- الخلية العصبية تستمد مميزاتا من الخلية العصبية الحيوية .  
- كل عصبون له Active معين ناتج من عملية المعالجة .  
- الروابط هي عبارة عن connection واوزان و sum وهي العملية الاساسية في العصبون ونعني بها المحاكاة للخلية العصبية الحية فالعصبون الصناعي هو محاكاة للعصبون الحي .  
امال sum هو جمع اشارة الدخل مضروبة في الاوزان  $\sum x_i w$

$x \Rightarrow$  cell /unit/node

$w \Rightarrow$  weight on connection

- العصبون في طبقة الدخل لا يتم عليه اي عملية معالجة بل ادخال فقط.

الشكل رقم (2) معمارية الشبكات متعددة الطبقات (Heikki .2008)



$t_k$  : تمثل مخرج الخلية .

$\delta_1$  : الخطأ .

تعديل الاوزان التي تربط الطبقة الخفية وطبقة المدخلات في الطبقة الثانية

$$w_{pq}^{new} = w_{pq}^{old} + a\delta_1\theta_q^2$$

$a = 0.5$  تمثل معامل التعلم .

6 - بعد الوصول الى الطبقة الخفية نعيد العمليات السابقة في تعديل الاوزان بعد تحديد قيمة الخطأ

$$\delta_2 = \theta_q^2 (1 - \theta_q^2) \sum_q w_{pq}^2 \delta_1$$

وبالتالي يتم تحديد الاوزان الجديدة بين طبقة المدخلات والطبقة الخفية

$$w_{kp}^{new} = old + a \delta_2 x_k$$

$w_{kp}$

تكرر هذه الخطوات لعدة مرات وذلك للحصول على اقل نسبة للخطأ .

### الاطار العملي

في هذا الفصل سوف يتم تطبيق ما تم التطرق اليه في الجانب النظري باستخدام بيانات تشمل انواع المؤشرات الكيماوية والفيزيائية المسببة لتلوث مياه الشرب ، حيث ان زيادة في نسبة المؤشرات يؤدي الى تلوث الماء، وقد تم استخدام تسع انواع من المؤشرات لكافة محافظات العراق عدا اقليم كردستان.

اما المقارنة فقد كانت بين الخوارزمية الجينية MWCD والشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية) Back Propagation

من حيث السرعة والدقة في ايجاد موقع الوسيط عندما يكون لدينا نموذج انحدار متعدد المتغيرات اللامعلمي .

وقد تم استخدام برنامج Matlab في التطبيق .

### - جمع البيانات

اعتمدت الدراسة على الاحصاءات البيئية للعراق لسنة 2013 الصادرة من الجهاز المركزي للاحصاء مقسمة الى (10) أشهر وتشمل المحافظات كافة عدا اقليم كردستان وقد تم استخدام (9) انواع من المؤشرات الكيماوية والفيزيائية المسببة لتلوث مياه الشرب موضحة في الجدول رقم (1) والذي يبين فيه المؤشرات الداخلة في مياه الشرب مع الرمز ووحدة القياس المطابقة للمواصفات العراقية اذ ان الزيادة في الحدود الدنيا للقياس تسبب تلوث في مياه الشرب .

اما البيانات المستخدمة في هذه الدراسة موضحة كالآتي :-

اما افضل خوارزمية مستخدمة في تطبيق الشبكات العصبية متعددة الطبقات هي خوارزمية الانبعاث الخلفي Back Propagation Algorithm يتم في هذه الخوارزمية ايجاد الخطأ بين المخرج المطلوب والفعل وتعديل الاوزان بالرجوع به من الطبقة الاخيرة الى الطبقات الخفية ثم الى طبقة المدخلات ، حيث يستخدم التعليم المراقب لتقليل الخطأ باتجاه الصفر ، خلال عملية التعليم يتم تدريب الشبكة باستخدام الدالة السيبيية ( الدالة التفعيلية) Activation functions والتي تكون قابلة للتفاضل

(Differentiation) بالاضافة ان لا تكون ثنائية القيمة وان تكون قادرة على النمذجة اللاخطية ، وتوجد انواع مختلفة من الدوال التفعيلية :-

1- الدالة الخطية ( Linear function ) وتستعمل عادتاً في خلايا طبقة المخرجات .

2- Sigmoid function تستعمل في الطبقات الاخرى . من خواص الانبعاث الخلفي (الارتدادية) تقليل الخطأ وقابليتها للتعامل مع البيانات المشوشة وكذلك قدرتها على التعامل مع الدوال الخطية واللاخطية القابلة للاشتقاق .

اما خطوات خوارزمية الانبعاث الخلفي (الارتدادية) هي كالآتي :-

- 1- اعطاء قيم عشوائية للاوزان  $w$  .
- 2- تعريف معامل التعليم ( قيمة معامل التعلم الدقيقة  $a=0.5$  ) .
- 3- ادخال مصفوفة تمثل مدخلات ومخرجات التي تم التدريب عليها .

4 - تطبيق عملية الانتشار الامامي لتحديد مخرجات الشبكة

$$z_p = \sum_k w_{kp}^1 x_k$$

$x_k$  : تمثل المدخلات التي تم التدريب عليها .

$w_{kp}^1$  : تمثل الاوزان العشوائية التي تربط طبقة المدخلات والعقد الخفية .

بعد مرور هذه القيمة على الخلية والممثلة في الدالة السيبيية sigmoid function يكون مخرج كل خلية في طبقة المخرجات

$$\frac{1}{1+e^{-z_p}}$$

$$\theta_p^1$$

$$y_q = \sum_p w_{pq}^2 \theta_p^1$$

$$\theta_q^2 = \frac{1}{1+e^{-y_q}}$$

5 - مقارنة المخرجات الفعلية مع المخرجات المطلوبة وتحديد قيمة الخطأ

$$\delta_1 = (t_k - \theta_q^2) \theta_p^2 (1 - \theta_q^2)$$



جدول رقم (1) المؤشرات الداخلة في مياه الشرب مع الرمز الحدود الدنيا للقياس ووحدة القياس للمواصفات العراقية

المؤشر	الرمز	وحدة القياس	الحدود الدنيا للقياس
العكورة	Turbidity	NTU	5
العسرة الكلية	T.H	mg/ L	500
القاعدية	ALK.	mg/ L	0.2
الأملاح الذائبة الكلية	T.D.S	mg/ L	1000
الكلوريدات	Cl	mg/ L	350
الكالسيوم	Ca	mg/ L	150
المغنسيوم	Mg	mg/ L	100
الصوديوم	Na	mg/ L	200
الكبريتات	SO4	mg/ L	400

المصدر: الاحصاءات البيئية للعراق لسنة 2013 الصادر من الجهاز المركزي للإحصاء .

(Fisher-consistency) ( $ch^+ = 0.0057$ ) لمستوى معنوي (0.5) وعدد المتغيرات (p=9) .

الخطوة الرابعة :

يتم ايجاد  $\hat{V}_2$  بعدها يتم ايجاد المسافة بالاستناد الى الوسط الحسابي  $\hat{M}_2$  الموزون ومصفوفة التباين المشترك  $\hat{V}_2$  وايجاد الرتبة المقابلة للمسافة ثم يتم ايجاد  $Q_2$  الموضحة في المعادلة (2-13) .

الخطوة الخامسة :

بعد اخذ التباديل لجميع العينات الجزئية حيث ان عدد التباديل 737 تتم المقارنة بين  $Q_1$  و بتطبيق الخوارزمية العامة c-step بعد ترتيب  $Q_1$  تصاعدياً لجميع التباديل يتم عند ذلك اختيار افضل عشرة التي تكون لديها اصغر دالة ( $Q_1$ ) ، ثم نحدد ( $\hat{M}, \hat{V}$ ) يكون لديها اقل قيمة للدالة ( $Q_1$ ) من بين القيم العشرة التي تم اختيارها و قيمتها ( $Q_1=8.4356e+03$ ) ، اما العينة الوسيطة

$$N=n_7+n_3+n_4+n_8+n_1+n_2+n_5+n_6+n_9 + n_{10}$$

#### - الشبكات العصبية متعددة الطبقات

لتطبيق الشبكات متعددة الطبقات لا بد من تحديد عدد الطبقات وعدد العصبون neuron لكل طبقة بالاضافة الى تحديد الاوزان التي سيتم توليدها والشكل التالي يوضح مخطط الشبكة التي سيتم التطبيق عليها :-

#### -Minimum Weighted Covariance Determinant (MWCD)

يتم في هذه الخوارزمية الاعتماد على الرتبة الى المسافة ، يتم فيها تحديد دالة الوزن الموضحة في المعادلة (2-5) كالآتي :-

$$an=0.499443826$$

$$\approx 0.5$$

المرحلة الاولى

الخطوة الاولى :

بعد تحديد الوزن يتم السحب العشوائي للعينات مع اخذ التباديل (permutations) لكل العينات ، يتم ايجاد المسافة (Md) للعينة الاولى بالاعتماد على الوسط الحسابي  $\hat{M}_1$  ومصفوفة التباين المشترك ثم يتم سحب عينة عشوائية اخرى واضافتها للعينة الاولى ، يتم ايجاد المسافة لكل عينة (لكل اضافة) وتكرر عملية السحب والاضافة لحين الحصول على (N=898) التي تمثل العينة الكلية .

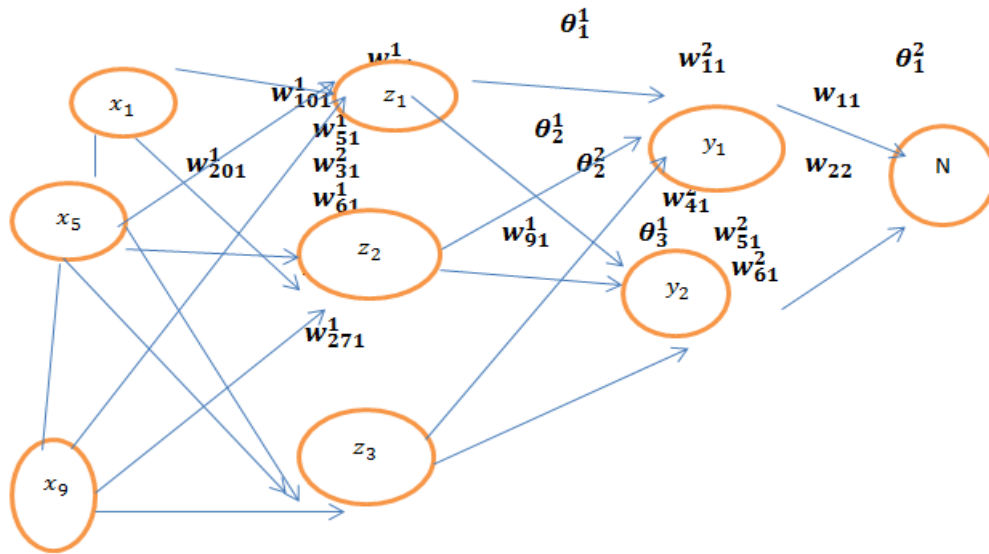
الخطوة الثانية :

يتم ترتيب المسافة تصاعدياً Md وايجاد الرتبة الى المسافة  $R_i$  ثم نجد  $Q_1$  الموضحة في المعادلة (2-6) .

الخطوة الثالثة :

بعدها يتم ايجاد الوسط الحسابي الموزون  $\hat{M}_2$  و  $\sum$  الموضحة في المعادلة (2-8) و (2-9) على التوالي ، اما قيمة العامل المتسق

الشكل رقم (3) تطبيق مخطط الشبكة



تم استعمال خوارزمية الانبعاث الخلفي (الارتدادية) والتعليم المراقب لتدريب الشبكة العصبية متعددة الطبقات ، ففي المرحلة الاولى تم توليد الاوزان العشوائية التي تدخل طبقة المدخلات وفيها يتم ضرب الاوزان مع كل عنصر دخل  $x_i$  ثم يتم ايجاد مجموع المدخلات للخلية الاولى في الطبقة الخفية كالآتي :-

$$z_p = \sum_k w_{kp}^1 x_i$$

يتم تمرير المدخلات وصولا الى طبقة المخرجات وذلك باستخدام الدالة السببية

$$\theta_p^1 = \frac{1}{1+e^{-z_p}}$$

$x_n$  : تمثل المدخلات (المؤشرات الداخلة في تلوث مياه الشرب) و  $n=1,2,\dots,9$

الطبقة الخفية و  $p=1,2,3$

الطبقة الخفية و  $q=1,2$

الاوزان العشوائية في الطبقة الاولى و  $k=1,2,\dots,27$

الاوزان العشوائية في الطبقة الثانية و  $j=1,2,\dots,6$

الاوزان  $w_{22}$  و  $w_{11}$  العشوائية

$\theta_p^1$  : العصبون في الطبقة الاولى الذي يربط طبقة المدخلات مع الطبقة الخفية

$\theta_q^2$  : العصبون في الطبقة الثانية

جدول رقم (2) جدول الاوزان العشوائية التي تم توليدها في الطبقة الاولى

الاوزان العشوائية في الطبقة الاولى
0.8234578283
0.6948286229
0.3170994800
0.9502220488
0.0344460805
0.4387443596
0.3815584570
0.7655167881
0.7951999011

0.1868726045
0.4897643957
0.4455862007
0.6463130101
0.7093648308
0.7546866819
0.2760250769
0.6797026768
0.6550980039
0.1626117351
0.1189976815
0.4983640519
0.9597439585
0.3403857266
0.5852677509
0.7546866819
0.7512670593
0.2550951154

يتم تمرير مدخلات الطبقة الخفية باستخدام الدالة السيبيية الاتية :

$$\theta_q^2 = \frac{1}{1+e^{-y_q}}$$

يتم بعدها توليد اوزان الطبقة الثانية حيث تضرب الاوزان مع المخرجات التي تم الحصول عليها من الطبقة الاولى وتجميعها في الطبقة الخفية كالآتي :-

$$y_q = \sum_p w_{pq}^2 \theta_p^1$$

جدول رقم (3) جدول الاوزان العشوائية في الطبقة الثانية

اوزان الطبقة الثانية
0.5645036389
0.7742749620
0.9494498398
1.0344896645
0.6057621172
0.2138226822

لتعديل الاوزان نبدأ اولاً بتحديد قيمة الخطأ  $\delta_1 = (t_k - O) * O * (1 - O)$  بعد تحديد قيمة الخطأ يتم تعديل الاوزان التي تربط بين الطبقة الخفية وطبقة المخرجات من المعادلة التالية :-

$$w_{11} \text{new} = w_{11} \text{old} + a \delta_1 \theta_1^2$$

$$w_{22} \text{new} = w_{22} \text{old} + a \delta_1 \theta_2^2$$

اما قيمة معامل التعلم  $a=0.5$

ثم الانتقال في هذه المرحلة من طبقة المخرجات الى الطبقة الخفية .

ولتعديل اوزان الطبقة الثانية يتم اولاً ايجاد قيمة الخطأ

يتم بعدها تحديد قيمة الخطأ وذلك من خلال مقارنة المخرجات الفعلية مع المخرجات المطلوبة ( $t_k$ ) اذا ان قيمة المخرجات الفعلية يتم ايجادها من ضرب الاوزان العشوائية  $w_{11}$  و  $w_{22}$  مع  $\theta_q^2$  وايجاد مجموع المدخلات لخلية طبقة المخرجات N

وبهذا يكون المخرج الفعلي للشبكة  $O = \frac{1}{1+e^{-N}}$  و

$$N = w_{11} \theta_1^2 + w_{22} \theta_2^2$$

اما قيمة الاوزان العشوائية التي تم توليدها هي كالآتي :-

$$w_{11} = 1.0062466544$$

$$w_{22} = 1.1013133923$$

$$\delta_3^1 = \theta_3^1(1-\theta_3^1)(w_{51}new\delta_{11} + w_{61}new\delta_{22})$$

اما الاوزان يمكن ايجادها من المعادلات التالية :-

$$w_{11}new = w_{11}old + a\delta_1^1 x_1 :$$

$$w_{91}new = w_{91}old + a\delta_1^1 x_9$$

$$w_{101}new = w_{101}old + a\delta_2^1 x_1 :$$

$$w_{181}new = w_{181}old + a\delta_2^1 x_9$$

$$w_{191}new = w_{191}old + a\delta_3^1 x_1 :$$

$$w_{271}new = w_{271}old + a\delta_3^1 x_9$$

تم تكرار خطوات خوارزمية الشبكة العصبية (50) مرة للحصول على اقل نسبة للخطأ بعدها يتم ايجاد مصفوفة التباين والتباين المشترك للعينات ثم تحديد افضل عينة من خلال ايجاد اصغر مسافة Mahalanobis Distance وقد تمثلت في مشاهدات العينة (n<sub>4</sub>) .

$$\delta_{11} = \theta_1^2(1-\theta_1^2)w_{11}new\delta_1$$

$$\delta_{22} = \theta_2^2(1-\theta_2^2)w_{22}new\delta_1$$

ثم تعدل الاوزان في الطبقة الثانية كالآتي :-

$$w_{11}^2 new = w_{11}^2 old + a\delta_{11}\theta_1^1$$

$$w_{21}^2 new = w_{21}^2 old + a\delta_{22}\theta_1^1$$

$$w_{31}^2 new = w_{31}^2 old + a\delta_{11}\theta_2^1$$

$$w_{41}^2 new = w_{41}^2 old + a\delta_{22}\theta_2^1$$

$$w_{51}^2 new = w_{51}^2 old + a\delta_{11}\theta_3^1$$

$$w_{61}^2 new = w_{61}^2 old + a\delta_{22}\theta_3^1$$

والان يمكن تعديل الاوزان في طبقة المدخلات بعد ايجاد قيمة الخطأ بين الطبقة الخفية الاولى وطبقة المدخلات كالآتي :-

$$\delta_1^1 = \theta_1^1(1-\theta_1^1)(w_{11}new\delta_{11} + w_{21}new\delta_{22})$$

$$\delta_2^1 = \theta_2^1(1-\theta_2^1)(w_{31}new\delta_{11} + w_{41}new\delta_{22})$$

جدول رقم (4) الجدول يوضح النتائج مبين فية التباين والمسافة والمحدد لافضل عينة محددة

المؤشرات	الشبكة العصبية Multilayer التباين	الخوارزمية الجينية MWCD التباين
Turb	48.2349	1.318e+9
T.H	9.1019e+03	1.0808e+12
ALK.	186.0992	1.3969e+11
T.D.S	3.5928e+05	3.9629e+12
CL	6.9147e+03	1.6613e+11
Ca	580.0939	6.2401e+10
Mg	264.4959	1.1083e+10
Na	1.9738e+03	5.0443e+10
SO4	7.1561e+03	3.6397e+11
المسافة (md)	6.4933e-26	8.4356e+03
المحدد (sd)	7.4450e+23	9.6518e+058

يشير الجدول اعلاه الى ان التقدير لموقع الوسيط باستخدام الشبكة العصبية الارجاعية (الارتدادية) افضل وذلك من خلال مقارنة التباين والمسافة والمحدد اقل بكثير من الخوارزمية الجينية MWCD .

$$y_i(t_j) = x_{r,i}(t_j)\beta_r(t_j) + \epsilon_i(t_j)$$

$$r = 1, 2, \dots, 9$$

$$i = 1, 2, \dots, 15$$

$$j = 1, 2, \dots, n_4$$

و ان :-

$$y_i(t_j) : 87*1$$

$$x_{r,i}(t_j) : 87*135$$

- تقدير معاملات الانموذج المتغيرة زمنيا  
في هذا المبحث سيتم تقدير معاملات الانموذج المتغيرة زمنيا" (بيانات القياسات المتكررة (repeated measurement) بايجاد معالم التمهيد لدالة المعاملات

$(\lambda_1 = 0.5672 , \lambda_2 = 1.9326 , \lambda_3 = 0.3284 , \lambda_4 = 0.4037 , \lambda_5 = 0.3035 , \lambda_6 = 0.3350 , \lambda_7 = 0.3009 , \lambda_8 = 0.3213 , \lambda_9 = 0.3063)$

$\beta_r(t_j)$  : متجه  $135 \times 1$

$\epsilon_i(t_j)$  : متجه  $135 \times 1$

ومن خلال كفاءة التقدير للشبكة العصبية الارجاعية (الارتدادية) التي تم مقارنتها مع الخوارزمية الجينية (MWCD) الممثلة بالعينة  $n_4$  ، لذا سيتم تقدير معاملات الانموذج للعينة  $n_4$  باستعمال شرائح التمهيد التكميلية والموضحة نتائجها في الجدول رقم (5) و (6) و (7) اما نتائج معالم التمهيد فهي كالآتي :-

رقم (5) تقدير معاملات العينة  $n_4$  لانموذج الشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية)

المحافظة	بغداد	صلاح الدين	الانبار	نينوى	ديالى
المعاملات					
$B_1(t_j)$	-0.1368	2.1029	3.4633	-1.4551	0.0671
$B_2(t_j)$	0.0194	-0.0074	0.0186	0.0124	0.0659
$B_3(t_j)$	2.0679e+08	-1.9640e+06	1.0215e+16	5.6548e+08	-6.1498e+09
$B_4(t_j)$	4.4688	503.9278	-359.9195	3.6856e+03	-152.965
$B_5(t_j)$	-6.2180e+29	-4.7957e+27	-1.0987e+26	-2.7939e+39	8.4498e+35
$B_6(t_j)$	-1.3882e+28	-5.7626e+29	-1.2624e+29	1.2600e+25	-1.7703e+27
$B_7(t_j)$	1.1806e+41	-3.3466e+38	3.9549e+38	6.7824e+45	-1.1895e+44
$B_8(t_j)$	-6.0566e+35	1.0358e+48	1.3925e+31	-3.6822e+49	4.5829e+22
$B_9(t_j)$	-7.6579e+16	-8.8761e+22	1.9365e+20	6.7159e+28	-8.7902e+20

الجدول رقم (6) تقدير معاملات العينة  $n_4$  لانموذج الشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية) .

المحافظة	بابل	ذي قار	كربلاء	النجف	المثنى
المعاملات					
$B_1(t_j)$	0.4447	0.1076	2.2284	0.8020	1.0677
$B_2(t_j)$	0.0029	0.0242	-0.0090	0.0088	0.0084
$B_3(t_j)$	-2.1315e+14	-8.4631e+08	-8.1258e+14	-7.0947e+19	1.6457e+12
$B_4(t_j)$	21.8447	-0.0667	-1.7809	-0.8350	0.0462
$B_5(t_j)$	-3.4744e+25	-3.2466e+21	4.1191e+26	-1.4570e+24	-4.4188e+17
$B_6(t_j)$	-4.6586e+27	-7.5678e+28	3.3603e+29	-3.7317e+26	-1.3052e+27
$B_7(t_j)$	-1.1859e+44	-5.8282e+38	1.9075e+41	-3.5059e+40	-8.3386e+45
$B_8(t_j)$	3.5592e+32	4.4087e+31	1.0717e+31	5.8199e+25	-3.8251e+24
$B_9(t_j)$	-6.5662e+13	-1.4949e+19	-2.6762e+16	2.3564e+16	-1.3322e+16

الجدول رقم (7) تقدير معاملات العينة  $n_4$  لانموذج الشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية).

المحافظة	كر كوك	القادسية	البصرة	ميسان	واسط
$B_1(t_j)$	-3.0743	0.0133	-7.5232	0.0162	1.4622
$B_2(t_j)$	0.0349	-0.0133	0.0080	-0.0448	-0.0040
$B_3(t_j)$	-1.0614e+06	-1.8191e+14	8.6728e+13	-2.9273e+15	-8.1519e+08
$B_4(t_j)$	-7.8626e+05	0.0397	-1.1918	4.2009	1.3392
$B_5(t_j)$	2.6791e+38	-2.7138e+26	4.7743e+16	3.9796e+16	6.5721e+23
$B_6(t_j)$	-1.4852e+27	3.3167e+29	-8.9767e+36	9.4747e+32	3.6500e+31
$B_7(t_j)$	-5.6134e+38	8.2124e+42	-4.5204e+44	4.5480e+38	-9.7334e+31
$B_8(t_j)$	4.9315e+47	-3.9929e+23	-1.3023e+37	1.2212e+28	-2.2046e+26
$B_9(t_j)$	1.4907e+34	1.9410e+19	-7.0228e+23	3.0293e+24	8.0225e+19

الحد الاعلى للقياس مما يدل على عدم تلوث المياه فيما يخص المؤشرين اعلاه لكل المحافظات .

3. بعد تحديد العينة الوسيطة وتقدير معاملات العينة  $n_4$  والتي يمكن من خلالها الاستدلال على تلوث المياه في بعض المحافظات والتي ظهرت في بعض من مؤشرات التلوث وذلك من خلال الحد الادنى للقياس المسموح به، حيث يمكن معالجة تلوثات المياه وتحديد المحافظة ونوع المؤشر المسبب للتلوث .

#### التوصيات

تطبيق الشبكة العصبية متعددة الطبقات عند زيادة حجم العينة في نموذج متعدد المتغيرات لسرعتها، وكذلك مقارنتها مع شبكات عصبية وكذلك خوارزميات جينية اخرى.

#### المصادر

هند رستم محمد شعبان، حسن ياسين طعمة، و حسن ثابت كرامشة. (2009). تحليل وتصميم الخوارزميات. 23 فبراير.

Azzini, A., Andrea G.B.Tettamanzi, & Vincenzopiuri. (2005). Anew Genetic Approach for Neural Network Design and optimization. UNIVERSITA DEGLISTUDI.

Brabanter, K. D., Johan A.K.suykens, & Bart De Moor. (2013). Nonparametric Regression Via stat Lssvm. journal of statistical software, 1-22.

Chakraborty, B., & Brobal Chaudhuri. (2005). On An optimization problem in robust. Retrieved from www2.ims.nus.sg/preprints/2005-26

يوضح الجدول رقم (5) و(6) و(7) تقارب في قيم المعاملات للعكورة وكذلك العسرة الكلية والذي يدل على عدم تلوث للمياه في كافة المحافظات ، ارتفاع القاعدية في المحافظات ( بغداد ، الانبار ، نينوى ، المثنى ، البصرة ) ، اما الاملاح الذائبة الكلية نلاحظ ارتفاع في قيم معاملاتها في محافظة ( نينوى ) عن الحد المسموح به ، وارتفاع الكلوريدات في المحافظات ( ديالى ، كربلاء ، كركوك ، البصرة ، ميسان ، واسط ) اما الكالسيوم فقد كانت نسبته مرتفعة عن الحد الادنى في المحافظات ( نينوى ، كربلاء ، القادسية ، ميسان ، واسط ) .

اما المغنيسيوم نلاحظ ارتفاعه في المحافظات ( بغداد ، الانبار ، نينوى ، كربلاء ، القادسية ، ميسان ) ، وارتفاع الصوديوم في المحافظات ( صلاح الدين ، الانبار ، ديالى ، بابل ، ذي قار ، كربلاء ، النجف ، كركوك ، ميسان ) ، اما ارتفاع الكبريتات كان في المحافظات ( الانبار ، نينوى ، النجف ، كركوك ، القادسية ، ميسان ، واسط ) ارتفاع قيم المعاملات عن الحد الادنى للقياس المسموح به يؤدي الى تلوث المياه .

#### الاستنتاجات والتوصيات

#### الاستنتاجات

1. انخفاض التباين باستعمال الشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية) وكذلك المحدد والمسافة مقارنتا بالخوارزمية الجينية MWCD ، سرعة الشبكة العصبية متعددة الطبقات الارجاعية (الارتدادية) اذ ان عدد تكرارات 50 في حين ان الخوارزمية الجينية MWCD عدد تكراراتها 737 حيث تم فيها اخذ التباديل للعينات الجزئية .
2. بعد تحديد العينة الوسيطة المتمثلة بالعينة  $n_4$  وتقدير قيم المعاملات لوحظ ان مؤشر العكورة والعسرة الكلية لم تتجاوز

- measurements. Journal of the American statistical Association, 1556-1569.
- WU, J. (1994). Neural Network and simulation methods. In I. Marcel Dakker. New york.
- Z.Huang, J., Colin O.wu, & Lan Zhou. (2004). Polynomial spline estimation and inference for varying coefficient model with longitudinal data. *statistica sinica*, 763-788.
- Zuo, Y. (2004). Projection-Based Affine Equivariant Multivariate Location Estimators with the Best Possible Finite sample Breakdown Point. *statistics sinica*, 1199-1208.
- D.crawford, K., & Rogerl.wainwright. (1995). Applying Genetic Algorithm to outlier Detection Morgan Kaufmn. Retrieved from Euler.mcs.utulsa.edu/~rogerw/.../Crawford-icga95.
- Despagne, F., & D.Luc massart. (1998). Neural Network in Multivariate Calibration. *Analyst*, 157-178.
- J.Rousseeuw, P., Stefan Van Aelst, Katrien Van Driessen, & Jose Agullo. (2003). Robust Multivariate Regression. Retrieved from Users.ugent.be/~svaelst/.../multi MCD-report.
- Johnston, R. (1996). Are speech Recognition still 98% Accurate or has the time come Repeal HYDE Law. *BT Tehnoly*, 165-176.
- Kinnerbrock, W. (1995). Neural Network. Suneel Galgotia.
- N.Koivo, H. (2008). NEURAL NETWORKS: Basics using MATLAB Neural Network Toolbox. pp. 1-59.
- P.P.B.Eggermont, R.L.Eubank, & V.N.Lariccia. (2005, May 18). Convergence Rates for smoothing spline Estimators in varying coefficient model. Retrieved from www.udel.edu/FREC/eggmont/.../Time Varying.
- piison, G., & Stefan Van Aelst. (2003). Retrieved from Diagnostic Plots for Robust Multivariate Methods: <https://lirias.kuleuven.be/..diagnostic-plots-report.pdf>
- R.Hoover, D., John A.Rice, Colin O.wu, & LI-PING YANG. (1998). Nonparametric smoothing estimates of time - varying coefficient models with longitudinal data. *Biometrika*, 809-822.
- Roelant, E., Stefan Van Aelst, & Gert Willems. (2009). The minimum weighted covariance determinant estimator. DOI1007/soo184-0186-3, 177-204.
- WANG, L., Hongzhe LI, & Jianhua UANG. (2008). Variable selection in Nonparametric varying - coefficient models for Analysis of Repeated