

استخدام الشبكات العصبية لأغراض التمييز Neural Networks as a Discriminant Purposes

م. م. طاهر ريسان دخيل
جامعة القادسية/ كلية الادارة والاقتصاد
قسم الاحصاء

الخلاصة

يعد اسلوب التمييز بين المجموعات من الاساليب الشائعة الاستخدام وذلك لكثرة الظواهر التطبيقية التي يمكن ان يتم تحليلها من خلال هذا الاسلوب، وهناك العديد من الطرائق التي يمكن ان تستخدم لهذا الغرض، مثل الدالة التمييزية الخطية، والتربيعية، وفي الفترة الاخيرة برز استخدام الشبكات العصبية *Neural Networks* كأداة للتمييز بين المجموعات . وفي بحثنا هذا تم استخدام المحاكاة لغرض المقارنة بين طريقة الشبكات العصبية والطريقة التقليدية لتصنيف المشاهدات الى المجموعة التي تنتمي اليها وذلك في حالة كون بعض المتغيرات تتوزع توزيعاً غير التوزيع الطبيعي وذلك لمعرفة افضلية الطريقتين المستخدمتين . وقد تم استخدام معيار نسبة المشاهدات المصنفة خطأ" الى مجموع المشاهدات الكلي كمعيار للمقارنة بين الطريقتين .

Abstract

Discriminant between groups is one of the common procedures because of its ability to analyze many practical phenomena, and there are several methods can be used for this purpose, such as linear and quadratic discriminant functions. recently, neural networks is used as a tool to distinguish between groups.

In this paper the simulation is used to compare neural networks and classical method for classify observations to group that is belong to, in case of some variables that don't follow the normal distribution. we use the proportion of number of misclassification observations to the all observations as a criterion of comparison.

المقدمة

برز في الآونة الاخيرة استخدام الشبكات العصبية *Neural Networks* لأغراض التنبؤ والسيطرة النوعية ومحاكاة عمل الرجل الالي وتحليل الصور والتمييز بين المجموعات وذلك بسبب المرونة التي يتمتع بها هذا الاسلوب رغم صعوبته وغموضه بالنسبة للكثير من الباحثين، وكذلك قدرته على التعامل مع الدوال غير الخطية بسبب عدم اعتماده على نوع التوزيع الذي تتبعه المتغيرات للظاهرة المدروسة، فهذا الاسلوب هو اسلوب تكراري يهدف الى جعل الخطأ اقل ما يمكن من خلال تعليم **Learning** ، او تدريب **Training** الشبكة العصبية بغية الوصول إلى الأوزان **Weights** التي تحقق المطلوب ((تقليل الخطأ)).

لقد كانت موجة الاهتمام بالشبكات العصبية بعد اعطاء الباحثان **McCulloch** و **Pitts** عام 1943 [3] مقدمة في الخلايا العصبية **Neurons** كخلايا بيولوجية، وكيفية محاكاة عمل هذه الخلايا لأغراض إجراء مهام حسابية، ولكن لم يتم استخدام الشبكات العصبية بكثرة في السنوات اللاحقة بسبب صعوبة تطبيقها وغموضها بالنسبة للباحثين، ولكن برزت موجة الاستخدام الأخرى بعد اقتراح اسلوب تدريب جديد يعتمد على الانتشار العكسي للخطأ من قبل الباحثين **Rumelhart** و **Hinton** و **Williams** [5] عام 1986 والذي أخذ مجالاً كبيراً في التطبيقات العملية لحد الآن

لقد استخدمت الشبكات العصبية لأغراض التمييز في العديد من البحوث، ففي عام 1990 وجد الباحثان **Surkan** و **Singleton** [6]، ان اسلوب الشبكات العصبية وتعطي نتائج أفضل من التحليل التمييزي التقليدي في التنبؤ المستقبلي لتقويم الخطورة في السنوات المالية . كذلك في عام 1991 قام الباحثان **Weiss** و **Kulikowski** [7] بتقديم طرقاً للتمييز مكونة من طرائق تقليدية وطرائق الشبكات العصبية ولأنواع مختلفة منها. وفي عام 1996 اعطا الباحثان **Krose** و **Smaget** [2] نظرة عامة عند استخدام الشبكات العصبية في التمييز مع اعطاء بعض اساليب التدريب لهذا الغرض . في عام 1997 قارن الباحثان **Jose** و **Moro** [1] بين طريقة التمييز التريبيعية وطريقة الشبكات العصبية وتوصلا الى ان طريقة الشبكات العصبية تعطي نتائج أفضل.

هدف البحث

يهدف هذا البحث الى استخدام الشبكات العصبية للتمييز بين مجموعتين عندما يتم خرق بعض الافتراضات للنموذج الرياضي لهذا الاسلوب، حيث تحتوي كل مجموعة على ستة متغيرات اربعة منها تتبع التوزيع الطبيعي واثان منها تتبع التوزيع اللوغارثمي الطبيعي وذلك عند احجام عينات مختلفة وباستخدام شبكة عصبية ذات طبقتين، احدهما الطبقة المخفية، التي تأخذ وحدات أو خلايا عصبية مختلفة والاخرى هي الطبقة المخرجة ومقارنة النتيجة التي يتم الحصول عليها مع الدالة التمييزية الخطية التقليدية من خلال استخدام معيار خطأ تصنيف المشاهدات الى مجموعتها الصحيحة لبيان افضلية الطريقتين في التمييز.

الدالة التمييزية الخطية (Linear Discriminant Function) [4]

تستخدم الدالة التمييزية الخطية عندما تكون المجتمعات المدروسة ذات توزيع طبيعي متعدد المتغيرات بمتجهات متوسط مختلفة ومصفوفة تباين وتباين مشترك متساوية .
أن عملية التمييز تتلخص بالخطوات التالية :-

1- إيجاد متوسط كل متغير في كل مجموعة ثم إيجاد الفرق بين متوسطي كل متغير في المجموعتين حيث أن متوسط المجموعة الأولى

$$\bar{X}_{(1)} = \sum_{j=1}^{n_1} X_j / n_1$$

متوسط المجموع الثانية

$$\bar{X}_{(2)} = \sum_{j=1}^{n_2} X_j / n_2$$

أما الفرق بين متوسطي المجموعتين d_i حيث أن

$$d_i = \bar{X}_{(1)} - \bar{X}_{(2)}$$

2- إيجاد مجموع المربعات لكل متغير في كل مجموعة وكذلك مجموع حاصل ضرب كل متغيرين داخل كل مجموعة أيضاً

$$s_{ii} = \sum X_j^2 - (\sum \bar{X}_i)^2 / n$$

$$s_{ij} = \sum X_i \cdot X_j - (\sum X_i)(\sum X_j) / n$$

3- إيجاد التباين والتباين المشترك المدموج (داخل المجاميع)

$$V_{ii} = \frac{S_{ii(1)} + S_{ii(2)}}{n_1 + n_2 - 2}$$

$$V_{ij} = \frac{S_{ij(1)} + S_{ij(2)}}{n_1 + n_2 - 2}$$

4- دالة التمييز تكتب كالاتي

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \dots + \alpha_r x_r$$

حيث $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_r$ يتم اختيارها بحيث تعطي أعلى تمييز بين المجموعتين

والمقصود بأعلى تمييز بين المجموعتين هو أن الاختلاف بين قيم المجموعتين **Between-Group variation** أكبر بكثير من الاختلافات بين قيم داخل المجموعتين **Within – Group Variation** فإذا رمزنا لنسبة الاختلافات بين المجموعتين إلى الاختلافات داخل المجموعتين بالرمز λ .

$$\lambda = \frac{\text{between – group. variation}}{\text{within – group. variation}}$$

فإننا نختار $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_r$ بحيث تكون λ أكبر مما يمكن .

5- إيجاد نقطة الفصل [4] Cutoff point

لتصنيف المشاهدات فإننا نحتاج إلى نقطة تفصل بين المجموعتين بحيث إذا قلت (مجموعة نقاط المشاهدات) عن تلك النقطة فإن المشاهدات عندئذ تتبع إلى مجموعة معينة وإذا زادت مجموع نقاطها عن تلك النقطة الفاصلة أصبحت تابعة للمجموعة الأخرى.

$$z = \frac{\bar{Y}_1 + \bar{Y}_2}{2}$$

حيث أن :-

$$\bar{y}_1 = \bar{x}(1) \cdot S^{-1}(\bar{x}(1) - \bar{x}(2))$$

$$\bar{y}_2 = \bar{x}(2) \cdot S^{-1}(\bar{x}(1) - \bar{x}(2))$$

قاعدة التصنيف [4] Classification Role

يمكن وضع قاعدة التصنيف وكالاتي:

$$y \geq Z$$

المشاهدة x تعود للمجتمع الأول إذا كان

$$y < Z$$

المشاهدة x تعود للمجتمع الثاني إذا كان

خطأ التصنيف [4] The Probability of Misclassification

هو احتمال تصنيف مشاهدة معينة إلى المجموعة الأولى بينما هي تعود في الحقيقة إلى المجموعة الثانية والعكس بالعكس نفترض لحساب خطأ التصنيف إن حجم العينة يكون كبير لذلك توزيع المشاهدات يقترب من التوزيع الطبيعي (حسب نظرية الحد المركزي) حيث يعتمد على أن العينة تتوزع توزيعاً طبيعياً أو يقترب من التوزيع الطبيعي.

$$P_{12} = \Phi(-1/2D)$$

حيث أن

$$D^2 = (\bar{X}_{(1)} - \bar{X}_{(2)})' * S^{-1} (\bar{X}_{(1)} - \bar{X}_{(2)})$$

ويتم إيجاد قيمة P_{12} من جداول التوزيع الطبيعي القياسي . أن خطأ التصنيف هو عامل مهم لإثبات كفاءة الدالة المميزة والدالة التي تعطي أقل خطأ تصنيف هي الدالة الأكثر كفاءة وتكون الأفضل من بين دوال التميز.

الشبكات العصبية الاصطناعية ANN [2][5]

إن الشبكات العصبية الاصطناعية ANN هي عبارة عن صيغ حسابية تعتمد على نماذج رياضية تحاكي عمل الدماغ Brain ، ويطلق على الشبكات العصبية أيضاً عدة مسميات أخرى مثل الأنظمة المترابطة Connection Systems ، او الانظمة التكيفية _Adaptive system، او انظمة التوزيع المتوازي Parallel Distributed Systems ، وان السر وراء هذه التسميات المتعددة هي بسبب ان الشبكات العصبية الاصطناعية مكونة من سلسلة من عناصر المعالجة وهي الخلايا العصبية Neurons والتي تكون مترابطة فيما بينها وفق خطوط متوازية ، بحيث ان جميع هذه العناصر تتغير وتحدث على وفق قواعد وطرائق تكيفية معينة والتي بدورها تعتمد على المعلومات المدخلة الى الشبكة العصبية الاصطناعية ANN .

إن من اهم اهداف الشبكات العصبية الاصطناعية ANN هي في فهم كيفية تكوين وتشكيل المميزات الدالية والخصائص الحسابية للدماغ عندما ينجز بعض العمليات مثل الإدراك الحسي والتمييز بين الأشياء والتعليم ومحاولة الاستفادة من هذه الخصائص لغرض تحويلها باتجاه رياضي وحسابي قادر على تحليل البيانات والمعلومات ومن ثم التنبؤ ، وفي الوقت الحاضر فإن هناك العديد من المحاولات التي تبذل في سبيل تقدير تطبيقات الشبكات العصبية الاصطناعية ANN واستخدامها في العديد من المجالات مثل تحليل السلاسل الزمنية والتحليل التمييزي وغيرها.

الشبكات العصبية كأساس للتصنيف [2] :

يمكن استخدام الشبكات العصبية الاصطناعية ANN كأساس للتصنيف وذلك من خلال تحديد دالة تنشيط *Activation function* ، خاصة بهذا الغرض، ودوال التنشيط هي عبارة عن دوال لمعالجة البيانات الداخلة الى الوحدات او الخلايا العصبية، وهناك العديد من دوال التنشيط التي قدمت من قبل الباحثين والتي اختلفت تبعاً لإختلاف نوعية المُخْرَج الذي يتم الحصول عليه واهداف الشبكة العصبية المراد تحقيقها، وهنا سوف يتم استخدام دالة الخطوة *Step function*، وتدعى ايضاً بدالة العتبة *threshold function* ، وذلك كونها مناسبة للإستخدامات التصنيفية أو التمييزية وذلك لانها تعطي قيمتين للمُخْرَج إما صفر أو واحد، وكما في الصيغة الآتية :

$$f(s) = \begin{cases} 0 & \text{if } s < 0 \\ 1 & \text{if } s \geq 0 \end{cases} \quad (1)$$

حيث تستخدم هذه الدالة في وحدات الطبقة المُخرَجة ، بينما سوف تستخدم الدالة اللوجستية Sigmoid في خلايا الطبقة المخفية والتي تعطي صيغتها بالشكل الآتي:

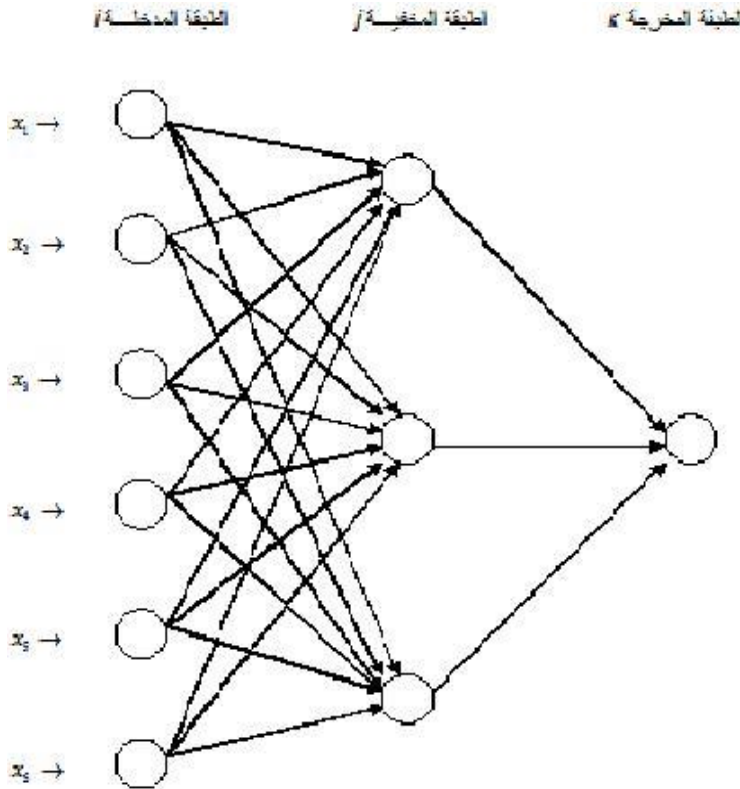
$$f(s) = \frac{1}{1 + e^{-s}} \quad \dots\dots\dots (2)$$

حيث ان s تمثل المجموع الموزون للمُدخلات مضافاً اليه حد التحيز (*Bias*) θ ، وفي هذا البحث تم استخدام متغيرات مُدخلة هي $(x_1, x_2, x_3, x_4, x_5, x_6)$ ، ضمن مجموعتين ، وبذلك فإن:

$$\dots\dots\dots (3) s = \sum_{i=1}^6 w_i x_i + \theta$$

إذن يمكن رسم الشبكة العصبية في حالة وجود ثلاث وحدات مخفية ووحدة واحدة مُخرَجة بالشكل

الآتي:



شكل (1)
يمثل شكل الشبكة العصبية للتجربة

حيث تحسب قيمة s في وحدات الطبقة المخفية على وفق الدالة اللوجستية، بينما في وحدة الطبقة المُخرَجة على وفق دالة الخطوة، وبذلك فإن مخرج الشبكة هو إما 0 أو 1، وذلك بالإعتماد

على المُدخل فإذا كان المُخرج 0 فهذا يعني ان المشاهدة تعود الى الصنف أو المجموعة الأولى، إما إذا كان المُخرج هو 1 فإن المشاهدة تعود الى الصنف أو المجموعة الثانية .

التدريب أو التعليم [6][5][2] Training or Learning

إن تقنية التدريب أو التعليم قد تم إستيحاء فكرتها من خلال قدرة الانسان على التعلم والتدرب من تجاربه التي يمر بها، وبهذا يمكن ان يعلم نفسه الى ان يصل الى حالة الاستقرار، ومن هنا فان تعليم الشبكة العصبية وتدريبها يعطيها القدرة على التعلم وبالتالي الوصول الى المُخرجات بأقل خطأ ممكن وذلك من خلال اعطاء الشبكة العصبية مجموعة من الامثلة والتي تدعى بفئة التدريب والتي تُحدث الاوزان بين الوحدات، من هذه الطرائق، طريقة تعليم الإدراك الحسي ((Perception)) ، وهذه الطريقة تُعدل وتُحدث الاوزان بأسلوب تكراري وذلك من خلال اضافة حد التعديل لقيمة الاوزان القديمة:

$$\left. \begin{aligned} w_i(t+1) &= w_i(t) + \Delta w_i(t) \\ \theta(t+1) &= \theta(t) + \Delta \theta(t) \end{aligned} \right\} \text{..... (4)}$$

وتكمن المشكلة هنا في كيفية حساب قيم $\Delta w_i(t)$ و $\Delta \theta(t)$ ، فعلى افتراض وجود مجموعة من عينات التعليم والذي يتضمن متجه المدخل x والمُخرج المرغوب به ((الحقيقي)) $d(x)$ ، حيث ان قيمة $d(x)$ في التصنيف أو التمييز هي اما صفر أو واحد، وبذلك يمكن توضيح طريقة تعليم الإدراك الحسي Perception من خلال الخطوات التالية :

1. إعطاء قيم ابتدائية للأوزان بين الوحدات .
2. ادخال متجه المدخل x من مجموعة عينات التدريب .
3. إذا كانت $y = f(x) \neq d(x)$ هذا يعني الحصول على استجابة غير صحيحة للتصنيف ، لذلك يتم تعديل الأوزان من خلال المعادلة الآتية:

$$\Delta \theta = d(x) \quad , \quad \Delta w_i = d(x)x_i$$

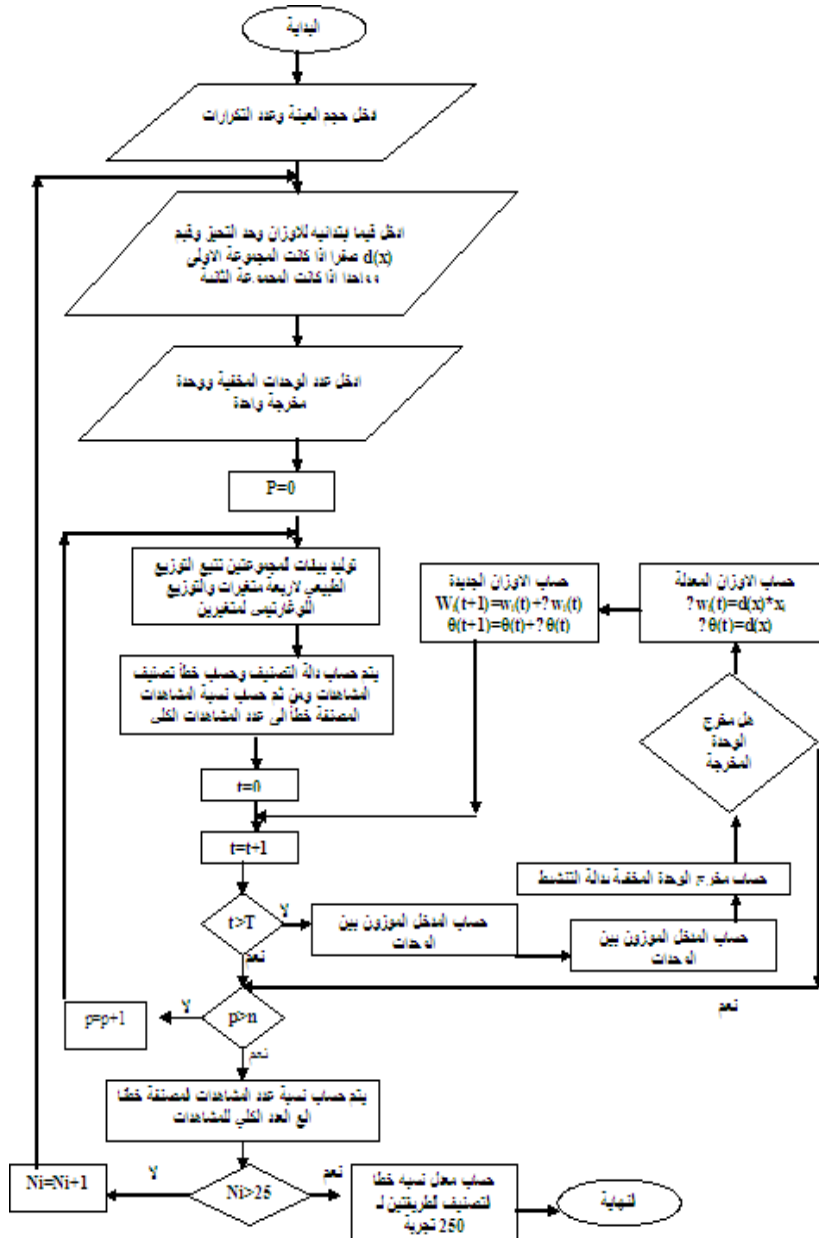
- إما إذا كانت $y = f(x) = d(x)$ هذا يعني الحصول على إستجابة صحيحة وبذلك فإن الأوزان تبقى نفسها وإن $\Delta \theta = 0$
4. أرجع الى الخطوة رقم (2)

تجربة المحاكاة

يمكن توضيح تجربة المحاكاة التي تم إرجاءها من خلال إستخدام برنامج Matlab بالخطوات الآتية:

1. تم افتراض أحجام للعينات وهي ($n=20,40,80,100,150$)
2. توليد n من لمشاهدات المتغيرات ($x_1, x_2, x_3, x_4, x_5, x_6$) وبمجموعتين ، المجموعة الاولى تأخذ القيمة (0) ، والمجموعة الثانية تأخذ القيمة (1) ، حيث تم حرق افتراض إن المتغيرات تتوزع توزيعاً طبيعياً، وذلك من خلال توليد متغيرين لا يتبعان هذا التوزيع، وانما يتبعان التوزيع اللوغارتمي الطبيعي.
3. تم افتراض عدد الوحدات المخفية 3,4,5,6 وحدة مخفية في الشبكة العصبية الاصطناعية، وبافتراض وجود وحدة مُخرجة واحدة .
4. يتم حساب الدالة التمييزية التقليدية، ومن ثم تصنيف المشاهدات على وفقها، وبالتالي حساب نسبة المشاهدات التي صنفت خطأً.
5. تم تدريب الشبكة العصبية بتكرار مساو الى 50 ، للوصول الى الأوزان الأفضل.
6. تم حساب نسبة المشاهدات التي صنفت خطأً من خلال الشبكة العصبية ومقارنتها مع نسبة المشاهدات التي صنفت خطأً في الخطوة رقم (4) .
7. تم اعادة تجربة المحاكاة 250 مرة لغرض الوصول الى نتائج مقنعة

والشكل التالي يوضح خوارزمية سير عملية المحاكاة:



شكل رقم (2)
يمثل خوارزمية سير عملية المحاكاة

تحليل النتائج

من خلال الجدول رقم (1) نلاحظ ان هناك أفضلية بصورة عامة لإسلوب الشبكات العصبية على طريقة التصنيف الخطية، حيث إن نسبة المشاهدات التي ارجعتها طريقة التصنيف الخطية هي أكبر من اسلوب الشبكات العصبية ولجميع أحجام العينات المستخدمة. نلاحظ كذلك انه عندما يكون عدد الوحدات في الطبقة المَخفية 4 ، تعطي نتائج أفضل عندما يكون حجم العينة كبير، وهو عكس ما نلاحظه عندما يكون حجم العينة صغيراً ومتوسطاً، حيث نلاحظ أن هناك ثبوتاً نسبياً في نسبة المشاهدات المصنفة خطأً.

جدول رقم (1)
يمثل نتائج تجربة المحاكاة

n	LDS	NN			
		node=3	node=4	node=5	node=6
20	0.184	0.161	0.161	0.162	0.161
40	0.182	0.158	0.158	0.155	0.154
80	0.182	0.159	0.155	0.156	0.153
100	0.171	0.140	0.139	0.143	0.143
150	0.172	0.141	0.140	0.144	0.144

الاستنتاجات

1. ان طريقة الشبكات العصبية تعطي نتائج أفضل من طريقة التصنيف الخطية.
2. عندما يكون حجم العينة كبير، يمكن الاعتماد على عدد الوحدات المخفية بأعداد مقارنة الى 3 أو 4 .
3. عندما يكون حجم العينة صغير أو متوسط، تكون النتائج متقاربة لأي عدد من الوحدات المخفية.

المصادر

- 1- Jose, I .& Moro, J . (1997) "comparing neural networks and multivariate discriminant analysis in the selection of new crop varieties", First European conference for information technology in Agriculture, Copenhagen , 15-18 June 1997
- 2- Krose, B. & smagt, p . (1996) "An introduction to neural networks", Eighth edition, The university of Amsterdam.
- 3- McCulloch, w. s. & pitts ,w . (1943) "A logical calculus of the ideas immanent in nervous activity" Bulletin of mathematical biophysics, 5, 115-133 .
- 4- Rencher, A . C . ((2002)) "Methods of Multivariate Analysis", second Edition , John Wily & sons, Inc
- 5- Rumelhart, D . E. , Hinton, G . E . & Williams , R . J . (1986) "learning representations by back . propagating error" , nature , 323 , 533 – 536 .
- 6- Surkar, A . & Singleton, J . ((1990)) "neural networks for Bond Rating Improved by Multiple Hidden Layers" proceedings of the 1990 International Joint conference on neural networks , 4 , 157 – 162 .
- 7- Weiss, s . m . & kulikowski, c . a . (1991) computer systems that learn " San Mateo : Morgan kaufimam .