جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل
كلية التربية للعلوم الصرفة
قسم الرياضيات

مقارنة بين استخدام الشبكات العصبية الاصطناعية والانحدار اللوجستي في التنبؤ

بحث مقدم إلى مجلس كلية التربية للعلوم الصرفة _ جامعة بابل
كجزء من متطلبات نيل شهادة البكالوريوس في الرياضيات

من قبل الطالبة
{ فاطمــه نوري عبيــس حمزه }
بإشراف
أ.م.د ايناس حمود محيسن

1444هـ 2023م

بسم الله الرحمن الرحيم

وقل رب زدني علما

صدق الله العظيم

سـورة طـه ايـة (١١٤)

الاهداء

الى مراد قلبي .. الأقرب الي من نفسي

المغيب عن الأبصار والكامن بعين البصيرة

الى والدنا المنتظر وأمامنا الحبيب صاحب العصر والزمان

قدوتي، ومثلي الأعلى في الحياة؛ فهو من علمني كيف أعيش بكرامة وشموخ فلقد كان له الفضل الأول في بلوغي إلى هذه المرحلة (والدي العزيز)

من علمتني العطاء، وغمرتني بحنانها وكرمها إلى من وضع

المولى سبحانه وتعالى الجنة تحت قدميها ووقرها في كتابه العزيز (أمي الحبيبة)

إلى جميع أساتذتي الكرام ممن لم يتوانوا في مد يد العون لي

أهدي إليكم بحث تخرجي

الشكـر والتقديـر

لا يسعني وأنا اضع اللمسات الأخيرة في هذا البحث إلا أن اتقدم بالشكر والعرفان لكل من ساعدني ودعمني واخص بالشكر والثناء الدكتورة إيناس حمود

التي كانت لها الفضل الكبير بالأشراف على هذا البحث وتزويدي

بالنصائح والارشادات والمعلومات من خبرتها الواسعة سائلا الله عز وجل أن يزيد في عملها ويرفع مقامها كما اتقدم بالشكر الجزيل لكل الأساتذة والدكاترة من الكادر التدريس

الذين لم يبخلوا علينا بتقديم المعلومات خلال المحاضرات الدراسية

سائلا المولى عز وجل أن يجعلها في ميزان حسناتهم

المحتويات

العنوان	الصفحة
الآية القرآنية	2
الاهداء	3
الشكر والتقدير	4
المحتويات	5
الملخص	6
الفصل الاول الشبكات العصبية الاصطناعية	16-7
(1-1) المقدمة	8
(2-1) تاريخ الشبكات العصبية الاصطناعية	9-8
(3-1) مفهوم الشبكات العصبية الاصطناعية	9
(4-1) مقارنة بين الخلية العصبية البشرية والخلية العصبية الاصطناعية	10-9
(5-1) مكونات الشبكات العصبية الاصطناعية	12-10
(6-1) عمل خوارزميات الشبكات العصبية	12
(7-1) كيف يتم التنبؤ بالشبكات العصبية	13-12
(8-1) طريقة بناء خوارزمية الشبكات العصبية	14-13
(9-1) العقد المخفية في الشبكات العصبية	14
(10-1) عدد طبقات الشبكات العصبية	15-14
(11-1) مزايا وعيوب الشبكات العصبية الاصطناعية	16-15
الفصل الثاني الانحدار اللوجستي	22-17
(1-2) المقدمة	18
(2-2) مفهوم الانحدار اللوجستي	20-18
(3-2) تحويلات الانحدار اللوجستي	20
(1-3-2) لوغاريتم معامل الترجيح	21-20
(2-3-2) لوغاريتم معامل الترجيح أو المفاضلة	21
(4-2) مزايا وعيوب الانحدار اللوجستي	22
الفصل الثالث مقارنة بين استخدام الشبكات العصبية الاصطناعية والانحدار اللوجستي	32-23
(1-3) التنبؤ	24
(2-3) الشبكات العصبية الاصطناعية	25-24
(1-2-3) النموذج الاحصائي	28-25
(3-3) مفهوم الانحدار اللوجستي	30-28
(1-3-3) مبررات استخدام الانحدار اللوجستي	31-30
(2-3-3) تقدير وتفسير معاملات الانحدار اللوجستي	32
المصادر	33

الملخص

تناول هذا البحث المقارنة بين أسلوب الشبكات العصبية وأسلوب الانحدار اللوجستي في التنبؤ حيث تم التعرف على الشبكات العصبية وهيه تقنيات حسابيه مصممه للمحاكاة مشابهة للطريقة التي يؤدي بها الدماغ البشري مهمه معينه عن طريق معالجه ضخمه موزعه على التوازي ومكونة من وحدات معالجه بسيطة تسمى عصبونات وأيضا تم التعرف على مفهوم الانحدار اللوجستي وهو أحد نماذج الانحدار التي تكون فيه العلاقة بين المتغير المعتمد (التابع) Y والمتغيرات المستقلة (المفسرة) X غير خطيه وغالبا ما تأخذ الاستجابة الشكل S ومن خلال هذا البحث توصل إلى أن أسلوب الشبكات العصبية الاصطناعية افضل من أسلوب الانحدار اللوجستي وباستخدام نموذج الشبكات العصبية الاصطناعية تم قياس درجة الأهمية للمتغير المستقل على المتغير التابع.

الفصل الاول

الشبكات العصبية الاصطناعية

1-1 المقدمة

تعتبر الشبكات العصبية من أهم مجالات الذكاء الصناعي الذي يعكس تطوراً هاماً ملموساً في طريقة التفكير الإنساني، وتدور فكرة الشبكات العصبية حول محاكاة العقل البشرى باستخدام الحاسب الآلي، وقد يعود التطور المنظور في هذا المجال إلى العديد من الدراسات التي تمت في مجال المعالجة العصبي (Neural Processing)وتتم عملية المحاكاة عن طريق حل المشاكل التي تواجهه، وذلك من خلال إتباع عمليات التعلم الذاتي والتي تعتمد على الخبرات المختزنة في الشبكة التي تحقق أفضل نتائج.

قام علماء الرياضيات بمحاكاة الشبكة العصبية في أبسط صورها برمجياً ، بصنع نموذج رياضي لها، يتميز بالسرعة العالية في معالجة البيانات كما يتميز بقدرته على التعلم والتعامل مع انماط مختلفة من البيانات التي قد تكون جزء منها خاطئاً مما جعلها مناسبة لكثير من التطبيقات مثل تمييز الصور والكلام، معالجة الإشارة، تمييز الصوت ...الخ

الشبكات العصبية الاصطناعية عبارة عن هيكل ذو بناء متوازي المعلومات، يتكون هذا الهيكل من وحدات معالجة تقوم بمعالجة المعلومات وتدعى بالعصبونات او عناصر الحساب، وتمر الإشارات بين العصبونات عبر خطوط ربط، وكل خلية عصبية تمثل ذاكرة محلية(Local Memory) كما يرفق كل خط بوزن (Weight) عددي معين يضرب مع الإشارات الداخلة للعصبون ثم يطبق على كل عصبون دالة تفعيل (Activation Function) على دخل الشبكة الذي يمثل مجموع اشارات الدخل الموزونة ليتم تحديد اشارة الخرج الناجمة عنه لها الفضل في حل العديد من المشاكل حيث دخلت في مجالات عديدة أهمها، مجال الطب: وهو تطبيق الطب الفوري الذي يرتبط بمبدأ الذاكرة كما في حالة العقل البشرى، أي مبدأ العالمات المرضية والتشخيص، و مجال الاتصالات السلكية واللاسلكية: مثل التخلص من صدى الذي قد ينتج في خطوط التليفونات، وفى الرادارات العسكرية لتحديد الأهداف، و مجال الأعمال المصرفية: وذلك لفتح الحسابات الخاصة بالبنوك عن طريق اللمس أو الصوت أو بصمة العين، وكذلك للتعرف على التوقيعات البنكية وخطوط اليد. و مجالات الأعمال: كتطبيق الشبكات في عدة أعمال وبصفة خاصة في مجال الأعمال الاقتصادية .

2-1 تأريخ الشبكات العصبية الاصطناعية

دوناك هيب في كتابه منظمه السلوك عام ١٩٤٩ أشار إلى أن المشابك العصبية الروابط (العصبونات) بين الخلايا العصبية تقوى كلما تم استعمالها أكثر، بمعنى أنه إذا كان يوجد عدد من الخلايا بجانب بعض ولكن اثنين منهم يقومان بنقل بيانات بصورة كثيفة فتقوى الوصلة بينهما، وتصبح عملية معالجتها للعمليات أسرع مع تكرار إثارتها بنفس المعطيات، كانت هذه بداية التفكير لما يسمى بالمعالجات العصبية أو الشبكات العصبية والتي كانت مطروحة في وقتها على صورة خاليا وليس شبكات مترابطة، في الخمسينات من القرن العشرين قامت شركه

(ABM) بأول محاولة لمحاكاة الخلية العصبية، ونجح ذلك بعد عدة محاولات فاشلة، ولكن كان علم الكمبيوتر في ذلك الوقت يتجه ناحية الحساب المتسلسل مما أدى إلى إهمال موضوع الخلايا العصبية ووضعه في الأدراج، و في عام 1959م قام (Bernard W.) ببناء نموذجي عنصر

تكيفي خطي آدا لاين Adaptive Linear Element ومجموعة عناصر تكيفية خطية Many ADA LINE . كان هذا هو أول ظهور للشبكات العصبية بشكلها الحالي، كانت تستخدم كفلاتر أو مرشحات قابلة (Adaptive Filter))لإلغاء الصدى من خطوط الهاتف، وما تزال تستعمل تجارياً حتى هذا الوقت، و استمرت الدراسات ولكن بشكل ضئيل بعد هذه الفترة إلى أن

قدم (Hopfield)بحثه (Neural Network and Physical Systems). والذي بين فيه إمكانية استخدام الشبكات العصبية في حل بعض المعضلات الصعبة المعروفة مثل)البائع المتجول(وبعض التطبيقات العملية، وتوالى التطور في علم الشبكات الاصطناعية إلى وقتنا هذا نتيجة للتطور المذهل الذي حصل في تقنيات الحواسيب الشخصية والحواسيب الصغيرة والتي تتمتع بقدرات حسابية فائقة وسعات خزن عالية فضال عن تطور البرمجيات وتوافر أدواتها مما ساعد على انتشار البحوث في هذا المجال وذلك من خلال إضافة أنواع جديدة من الشبكات وابتكار خوارزميات جديدة لتعليم الشبكات العصبية وزيادة كفاءته.

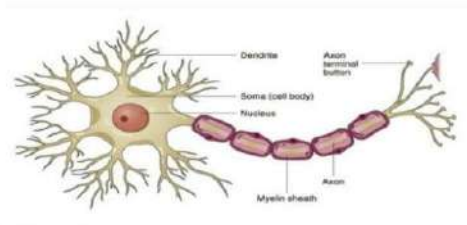
1-3 مفهوم الشبكات العصبية الاصطناعية

هي تقنيات حسابية مصممة لمحاكاة الطريقة التي يؤدي بها الدماغ البشري مهمة معينة، وذلك عن طريق معالجة ضخمة موزعة على التوازي، ومكونة من وحدات معالجة بسيطة، هذه الوحدات ما هي إلا عناصر حسابية تسمى عصبونات أو عقد (Neurons،Nodes) والتي لها خاصية عصبية ، من حيث أنها تقوم بتخزين المعرفة العملية والمعلومات التجريبية لتجعلها متاحة للمستخدم وذلك عن طريق ضبط الأوزان.

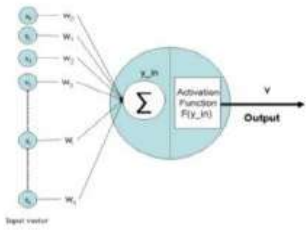
إذاً الANN تتشابه مع الدماغ البشري في أنها تكتسب المعرفة بالتدريب وتخزن هذه المعرفة باستخدام قوى وصل داخل العصبونات تسمى الأوزان التشابكية. وهناك أيضا تشابه عصبي حيوي مما يعطي الفرصة لعلماء البيولوجيا في الاعتماد على ANN لفهم تطور الظواهر الحيوية تعكس الشبكات العصبية سلوك العقل البشري، مما يسمح لبرامج الحاسب الآلي بالتعرف على النماذج وحل المشاكل الشائعة في مجالات الذكاء الاصطناعي والتعلم الآلي والتعلم العميق.

1-4 مقارنه بين الخلية العصبية البشرية والخلية العصبية الاصطناعية

كما رأينا قد قام العلماء باقتراح بناء نظام يحاكي العملية الموجودة في الخلية العصبية عند الانسان فلو نظرنا إلى الشكل (1-1) و الشكل (1-2) :



الشكل(1-1) خلية عصبية بشرية.



الشكل (1-2) خلية عصبيه اصطناعية.

نجد أنه يتكون من المدخلات (input vector) و هنا تمثل بال (X1 X3 X) . ويمكن أن تتخيل أنها تمثل التفرعات العصبية Dendrites للعصبون البشري والتي يتم من خلالها نقل السيالة العصبية من أعضاء الحس إلى جسم الخلية العصبية أي مجموع الإشارات المدخلة للخلية، وهنا تكون إما يوجد إشاره أي (واحد) أو لا يوجد إشارة (صفر).

الوزن (Weights) :ويمثل درجه الوزن للأشارة المدخلة حيث يعبر الوزن عن شدة الترابط بين عنصر قبله وعنصر بعده) .

تابع التنشيط (Activation Function) : وهنا يكمن العمل الحقيقي للخلية العصبية.

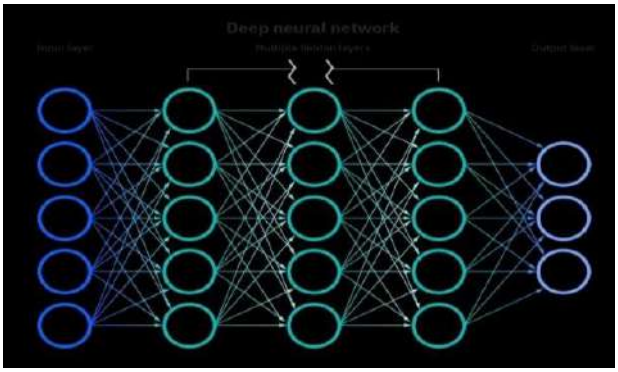
أي مثلا هنا يتم جمع الأوزان للإشارات المدخلة ومقارنتها بقيمة معينة للحد أو العنية (Threshold) فإذا كان مجموع أوزان الإشارات يزيد عن ال Threshold تكون الإشارة المخرجة هي (واحد) و اذا كان أقل يكون الناتج (صفر).

1-5 مكونات الشبكات العصبية

كما رأينا أن الشبكات العصبية تتكون من مجموعة من وحدات المعالجة ويسمى أحدها عصبون

كما أن للإنسان وحدات إدخال توصله بالعالم الخارجي وهي حواسه الخمس، فكذلك الشبكات العصبية تحتاج لوحدات إدخال . ووحدات معالجة يتم فيها عمليات حسابية تضبط بها الأوزان و نحصل من خلالها على ردة الفعل المناسبة لكل مدخل من المدخلات للشبكة . فوحدات الإدخال تكوّن طبقة تسمى طبقة المدخلات، و وحدات المعالجة تكوّن طبقة المعالجة وهي التي تخرج نواتج الشبكة. وبين كل طبقة من هذه الطبقات هناك طبقة من الوصلات البينية التي تربط كل طبقة بالطبقة التي تليها والتي يتم فيها ضبط الأوزان الخاصة بكل وصلة بينية، وتحتوي الشبكة

على طبقة واحدة فقط من وحدات الإدخال ، ولكنها قد تحتوي على أكثر من طبقة من طبقات المعالجة. والشكل (1–3) يوضح مكونات الشبكة العصبية.



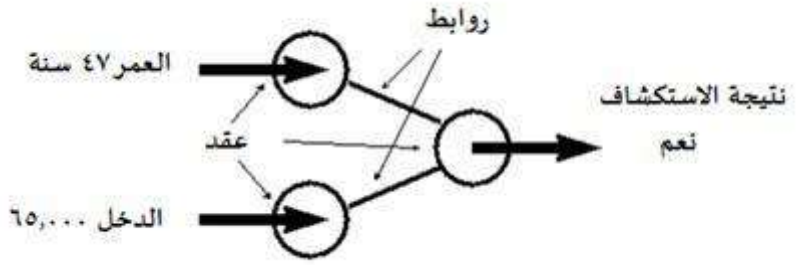
الشكل (1–3) يوضح مكونات الشبكات العصبية الاصطناعية

تعتمد الشبكات العصبية على بيانات التدرب لكي تتعلم وتحسن من دقتها بمرور الوقت. ومع ذلك، بمجرد ضبط خوارزميات التعلم هذه بدقة، فهي تعد أدوات قوية في علوم الحاسب والذكاء الاصطناعي، مما يسمح لنا بتصنيف البيانات وتجميعها بسرعة عالية. قد تستغرق المهام الخاصة بإمكانية التعرف على الكلام أو إمكانية التعرف على الصور دقائق في مقابل ساعات عند مقارنتها بالتعريف اليدوي من قبل الخبراء البشريين. واحدة من أكثر الشبكات العصبية المعروفة هي خوارزمية Google للبحث. تعتبر خوارزميات الشبكات العصبية أو (بالإنجليزية: Neural Networks) هي وأشجار القرار من أهم خوارزميات تنقيب البيانات. وذلك نظرًا للنتائج الدقيقة التي يتم التوصل إليها باستخدام هذه الخوارزميات. وكذلك لإمكانية تطبيقهما في حل العديد من المشاكل وبكافة الأنواع. هذا بالرغم من صعوبتهما مقارنة بالخوارزميات الأخرى والتي أدت لعدم الانتشار بشكل واسع لهما. وخوارزميات الشبكات العصبية تشبه في تركيبتها تركيبة عقل الإنسان. فهي تعمل بنفس الطريقة كما يعمل العقل في نقل ومعالجة وتحليل المعلومات والتوصل إلى الاستنتاجات واكتشاف الأنماط والتنبؤات. ونستطيع من خلالها تطبيق بعض ما يطبقه العقل الطبيعي، رغم أن العلماء لا يزالون حتى اليوم يكتشفون المزيد عنه ولم يلموا بكل تفاصيل عمله. وقد ساهمت طبيعة خوارزميات هذه الشبكات في أن تكون هي الأكثر استخدامًا في مجال الذكاء الاصطناعي. وذلك باعتبار أنها تهدف إلى محاكاة الذكاء البشري وإكساب الآلة بعض قدرات العقل الطبيعي. ويتم بناء خوارزميات الشبكات العصبية بطرق معقدة جدًا ومبنية على أنواع البيانات الرقمية. وعادة يتم استبدال قيم المتغيرات الاسمية المتوفرة بقيم رقمية لكي يتم استخدامها في الشبكة. وذلك من خلال عمليات تحويل البيانات وتفريد البيانات التي تتم

في مرحلة تحضير البيانات للتحليل والتنقيب. وكثيرًا ما تصب جهود المحللين باتجاه تبسيط وتسهيل خوارزميات من هذا النوع.

1-6 عمل خوارزميات الشبكات العصبية

إن خوارزميات الشبكات العصبية تعمل بنفس الطريقة التي يعمل بها مخ الإنسان. فهناك العُقد (التي تناظر الخلايا العصبية). وهناك الروابط التي تصل بينها (التي تناظر الوصلات العصبية). الشكل (1-4) يوضح تركيب شبكة عصبية بسيطة:



الشكل (1-4) تركيب شبكة عصبية بسيطة.

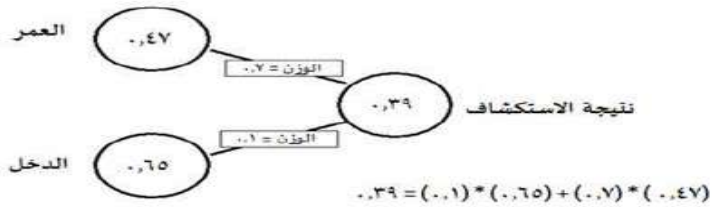
وفي هذا المثال، الشبكة تأخذ متغيري السن أو العمر والدخل وتعطي نتيجة تنبؤيه. وهدفها استكشاف ما إذا كان الشخص سيُقبل على عرض معين لشراء أحد المنتجات مثل شراء أحد أجهزة الحاسوب.

1-7 كيف يتم التنبؤ في الشبكات العصبية:

لكي يتم التوقع أو التنبؤ باستخدام الشبكات العصبية، يتم إدخال قيم المتغيرات المعلومة في العقد المخصصة للإدخال. ويصبح لكل عقدة قيمة السمة أو المتغير الذي تم إدخاله.

بعد ذلك يتم ضرب قيمة كل عقدة بقيمة الرابط المتصل بها وتجميع كل النتائج بحسب المعادلة الرياضية المعرفة في الشبكة العصبية ومن ثم الحصول على نتيجة التنبؤ بوقوع الحدث الذي يتم استكشافه. وفي هذا المثال، تم اعتبار أنه إذا كانت النتيجة (صفر) يكون من المتوقع استجابة

الشخص للعرض المقدم لـه. وإذا كانت النتيجة مقدارها (1) فيكون من المتوقع عدم استجابته ولتبسيط هذه العملية يمكن وصفها بالشكل (1–5):



الشكل (1–5) التنبؤ في الشبكات العصبية.

شبكة عصبية تأخذ متغيري العمر والدخل للشخص وتتنبأ بإقباله على شراء منتج معين – خوارزميات الشبكات العصبية وقد تم التعبير عن العمر والدخل بقيم تقع على التدريج بين (صفر، 1). فالعمر في هذا المثال 47 سنة وهو يناظر القيمة 0.47 على التدريج.

والدخل في هذا المثال بقيمة 65,000 والمناظر للقيمة 0.65 على المقياس أو التدريج.

أما الوصلات والتي تعبر عن الأوزان فقد تم تقديرها بالقيم (0.7) و(0.1) على الترتيب أو التتالي. وبعد ضرب قيم العقد في قيم الوصلات ثم جمعها نحصل على قيمة المتغير الذي نريد التوقع له فيكون الناتج هنا (0.39). وهي قيمة أقرب للقيمة (صفر) منها للقيمة (1).

وبذلك يتم اعتبارها (صفر)، وتكون بالتالي النتيجة هي توقع استجابة الشخص للعرض المقدم لـه لشراء جهاز الحاسوب.

1–8 طريقة بناء خوارزميات الشبكات العصبية

يتم بناء خوارزميات الشبكات العصبية عن طريق دراسة وتحليل أحداث وقعت بالفعل قبل ذلك ومخزنة في سجلات سابقة في قواعد البيانات.ثم بمقارنة التوقعات الناتجة من الشبكة العصبية مع السجلات الحقيقية السابقة يتم تطوير وتحديث قيم الأوزان في الشبكة بقيم أكثر دقة. وهذه يشبه إلى حد كبير عملية تصحيح أوراق إجابات الطلاب في المدرسة، فكلما كانت الأخطاء كبيرة فإنه يلزم لتصحيحها جهد أكبر. وكلما كانت الأخطاء قليلة يقل الجهد المطلوب لتصحيحها. وهكذا يتم بناء الشبكات العصبية بشكل تدريجي لتصبح أكثر صحة ودقة وكفاءة في تقدير التوقعات. ويعتبر تحديد قيم الأوزان، المعبر عنها بالوصلات، هي الأهم لبناء شبكة

عصبية يتم باستخدامها التوقع أو التنبؤ بشكل جيد ودقيق. وفي المثال السابق كانت الشبكة العصبية بسيطة جدًا وتهدف للتوضيح. في الواقع، يمكن أن تكون خوارزمية الشبكة العصبية أكثر تعقيدًا وتحتوي على العديد من العقد والوصلات التي قد تصل إلى عدة مئات.

1-9 العقد المخفية في الشبكات العصبية

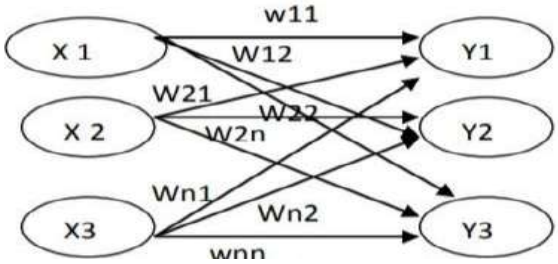
قد تحتوي خوارزميات الشبكات العصبية على نوعية أخرى من العقد والتي تسمى العقد المخفية. وتكون مهمة هذه العُقد استشارية ولا يؤخذ بقيمها إلا بعد أن يتم اعتماد استشارتها في حالة صحتها وبعد التجربة الفعلية وليس قبل. فالعُقد المخفية هنا تلعب دورًا استشاريًا فقط. ومثلما يحدث في الجيش، فالقائد يستمع إلى العديد من الاستشارات ممن حوله من المستشارين قبل اتخاذ قرار معين. ولكنه بعد اتخاذ القرار واستكشاف نتائجه ومدى صحته، يصبح بإمكانه تمييز المستشارين الجيدين والذين كانت آرائهم أقرب للقرار الذي كان من المُفترض أن يكون أنسب.

وبالتالي سوف يعتمد آرائهم في المستقبل ويأخذ بها أكثر من آراء غيرهم. وهكذا، فالعقد المخفية تلعب نفس هذا الدور. كلما تم تطبيق الخوارزمية يتم تطوير وتحديث العقد الأصلية. وذلك بأن تأخذ بالاعتبار قيم العقد المخفية المناسبة والتي تدعم الحصول على نتائج أكثر دقة. وبالمقابل يتم إهمال قيم العقد المخفية التي لم تحقق ذلك.

1-10 عدد طبقات الشبكة العصبية

1. شبكات الطبقة المنفردة

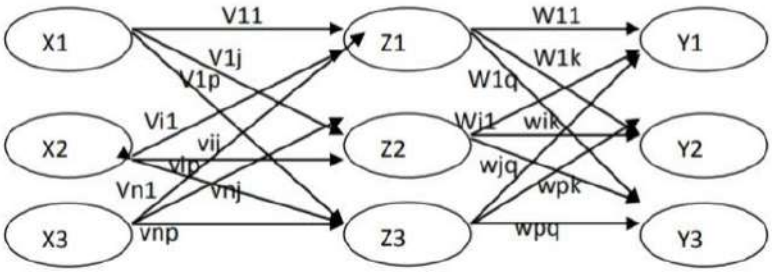
تمتلك هذه الشبكة طبقة واحدة من الأوزان إما وحدات الإدخال فيمكن تعريفها بأنها الإشارات التي تســتلمها من العالم الخارجي . وترتبط هذه الوحدات بوحدات الإخراج (التي تمثل الإجابة لوحدات الإدخال) وذات ارتباط كامل ومن أمثلتها شبكة (auto hetor Adeline preceptor) ذات طابع انســياب الخطأ خلفا ولا ترتبط وحدات الاخراج بأي وحدات أخرى باستثناء شبكه (Hopfiald) التي فيها كل وحدات الإدخال هي نفســها وحدات الإخراج والشكل (1-6) يمثل نموذج لهذه الشبكات



الشكل (1-6) شبكات الطبقة المنفردة

2. الشبكات متعددة الطبقات

هي الشبكات التي تمتلك طبقة أو أكثر من العقد المخفية (Hidden nodes) يمكن تعريفها بانها الوحدات التي تمثل وحدات إدخال و وحدات إخراج يمكن لهذه الشبكة التي تحل مشاكل أكثر تعقيدا من النوع الأول ولكن تدريبها يكون أصعب من أمثلتها (necogmatron miff , Back propagation Medline). الشكل (1-7) يمثل نموذج هذه الشبكات .



الشكل (1-7) يمثل معمارية الشبكات متعددة الطبقات

وفي الشبكة العصبية ثلاث طبقات من الأوزان هي :

- طبقة الوزن بين مستوى الإدخال والمستوى المخفي.
- طبقات الوزن بين المستويات المخفية.
- طبقة الوزن بين المستوى المخفي ومستوى الإخراج.

1-10 مزايا وعيوب الشبكات العصبية

أولاً: مزايا الشبكات العصبية

يتميز أسلوب الشبكات العصبية عن غيره من البدائل الإحصائية في عملية تحليل البيانات بالمزايا التالية:

- الموازاة في المعالجة: إن هندسة الشبكة مبنية على هذا الأساس إذ أنها تقوم بتركيب معقد لكل المتغيرات المستعملة ومعالجتها في نفس الوقت مما يقلص وقت التنفيذ، مقارنة مع طريقة البرمجة الكلاسيكية التي تعتمد على المعالجة التسلسلية للمتغيرات.
- القدرة على التكيف: إن قدرة التمرن الفعالة للشبكة تؤهلها الاستيعاب محددات جديدة للمشكلة من خلال المعطيات الجديدة للمحيط الخارجي.
- الذاكرة الموزعة: يتم تمثيل الحدث في تقنية الشبكات العصبية الصورية، حسب خريطة تنشيط للخلايا مما يسمح بالتصدي للصخب، زيادة على أن ضياع عنصر من الشبكة لا يؤثر على أدائها.
- القدرة على التعميم: من خلال عملية التمرن تتمكن الشبكة من التعرف على الأمثلة المشابهة.
- سهولة بناء نموذج الشبكة الاصطناعية: ويتم ذلك بكتابة برنامج والقيام بالاختيار للمعطيات وهذا ال يحتاج لوسائل كبيرة.
- بمجرد أن تقوم الشبكة بتلقي البيانات فإن أدائها لا يتأثر كثيراً عند ادخال اي مجموعة اضافيه البيانات الجديدة والتي لم تكن موجودة في البداية .
- لا تتطلب توافر عدد كبير من البيانات التجريبية.
- ليس هناك حاجة لترتيب المتغيرات حسب أهميتها، وذلك حيث أن الشبكة تضع تلقائياً أوزان خاصة بها حسب طبيعة البيانات المدخلة إليها.

ثانياً: عيوب الشبكات العصبية

و بالرغم من كل هذه الميزات للشبكات العصبية الا انه يوجد لها بعض العيوب وتتمثل في:

- من الناحية التقنية: عدم استغلال خاصية الموازاة في المعالجة إذ أن المحاكاة تتم حالياً على أجهزة ذات معالجة تسلسلية كلاسيكية مما يستغرق الكثير من الوقت.
- يجب مراعاة الاختيار السليم للمعطيات والترميز السليم لها والتشخيص الصحيح للظاهرة و أيضاً عملية المدخلات والمخرجات للوصول إلى نموذج فعال.
- ويعاب عليها أيضاً أنها لا تتعامل مع المشاكل ذات المتغيرات التي ال يمكن صياغتها كمياً كما أنها تعطي نتائج في بعض الحالات قد ال نستطيع أن نقدم تفسيراً منطقياً لها.

الفصـــل الثانـــي

الانحدار اللوجستي

2-1 المقدمة

من المألوف في دارسات الإنسانية والاجتماعية والاقتصادية ان يكون المتغير التابع متغيراً منفصلاً (نوعياً) بحيث يأخذ قيمه ثنائيه Dichotomous أو أكثر ، وان هذا يشكل تحدياً كبيراً للباحثين عند محاولتهم توظيف تحليل الانحدار الخطي (البسيط او المتعدد)، الذي يكو ن مقيد نوعا باشتراط ان يكون المتغير التابع متغير او كميا متصلا بدال من ان يكون وصفيا منفصلا. كما يذكر بانه يجب استخدام تقنية الانحدار اللوجستي Analysis Logistic

Technique في مثل هذه الحالات ، وان كانت هناك العديد من الأساليب الإحصائية التي تطورت لتحليل البيانات ذات المتغيرات الوصفية قوس(النوعية) مثل تحليل الدوال التمييزية Discriminant Functions Analysis، الا أن الانحدار اللوجستي يتميز بالعديد من المميزات التي تجعله ملائما للاستخدام في مثل تلك الحالات كما يتميز الانحدار اللوجستي في كونه أكثر مرونة من نماذج الانحدار التقليدية، اذ بالإمكان افتراض علاقه معينة تربط بين المتغير التابع والمتغيرات التوضيحية (المستقلة) الأخرى، ومن خلاله نستطيع مباشره تقدير احتمال حدوث حدث ما وتمكن اهميه تحليل الانحدار اللوجستي عند مقارنته بالأساليب الإحصائية الأخرى (الانحدار الخطي والتحليل التمييزي)، في أن الانحدار اللوجستي هو أداة أكثر قوة لأنه يقدم اختباراً لمعنوية المعاملات، كما انه يعطي الباحث فكرة عن مقدار تأثير المتغير المستقل في المتغير التابع النوعي الثنائي القيمة، بالإضافة الى ذلك، فأن الانحدار اللوجستي يرتب تأثير المتغيرات المستقلة، مما يسمح للباحث الاستنتاج بأن متغيرا ما يعتبر أقوي من المتغير الآخر في فهم ظهور النتيجة المطلوبة، كما ان تحليل الانحدار اللوجستي يمكنه ان يتضمن المتغيرات المستقلة النوعية، و هناك نوع آخر وهو الانحدار اللوجستي المتعدد Regression Multinomial Logistic المستخدم في حالة المتغير التابع الاسمي متعدد القيم (اكثر من قيمتين)، كما ان هناك نوع ثالث للانحدار اللوجستي يسمي الانحدار اللوجستي الرتبوي (الترتيبي). Ordinal Logistic Regressionالذي يستخدم في الحالات التي يكون فيها المتغير التابع متغير رتبوي علما أننا سنستخدم ترميز المتغير التابع ثنائي القيمة بالقيمتين(0،1) دون غيرها من اشكال الترميز الأخرى في هذه الدراسة.

2-2 مفهوم الانحدار اللوجستي:

الانحدار اللوجستي هو أحد نماذج الانحدار التي تكون فيها العلاقة بين المتغير المعتمد

(التابع) Yو المتغيرات المستقلة(المفسرة)X غير خطية، وغالبا ما تأخذ دالة الاستجابة للنموذج

الشكل s ورياضيا تعرف كالآتي:

$$p_i = E(Y_i \backslash X_i) = \frac{e(B_o + B_1X_1 + B_2X_2 + \cdots + B_kX_k)}{1 + e(B_o + B_1X_1 + B_2X_2 + \cdots + B_kX_k)} \quad (1.2)$$

$$i = 0,1,2, \ldots \ldots, k$$

أن نموذج اللوجستي أكثر مرونة من نماذج الانحدار التقليدية، إذ بالإمكان افتراض علاقة معينه تربط بين المتغير التابع والمتغيرات المفسرة الأخرى، ومن خلاله نستطيع تقدير احتمال حدوث حدث ما فضال عن سهولة تحو يله الي الشكل الخطي باستخدام ما يعرف بتحويل اللوجيت (logit trans formation) والمعروف كالاتي :

$$logit(p_i) = In\left[\frac{p_i}{1 - p_i}\right] = B_o + B_1X_1 + B_2X_2 + \cdots + B_kX_k \qquad (2.2)$$

حيث أن :

$[\frac{p_i}{1-p_i}]$: تمثل ما يسمى بنسبة الرجحان وهي نسبة احتمال حدوث الحدث الى عدم حدوث ذاك الحدث.

$logit(p_i)$: تمثل اللوغاريتم الطبيعي لنسبة الرجحان.

B_o:تمثل الجزء المقطوع (ثابت المعادلة).

B_i: المعامل اللوجستي لكل حاله من حالات المتغير المستقل حيث i=1,2,3,……,k

X_i: المتغيرات المستقلة حيث i=1,2,3,…,k

حيث اشار أن الانحدار اللوجستي هو أسلوب إحصائي لفهم العالقة بين المتغير التابع ذو المستوي الو صفي ومتغير واحد أو أكثر من المتغيرات المستقلة، و التي تسمي أحيانا متغيرات مصاحبه أو متغيرات مفسرة بحيث تكون تلك المتغيرات المستقلة من أي نوع من مستويات القياس. و نموذج الانحدار اللوجستي نموذج يستخدم للتنبؤ باحتمالية وقوع حدث ما وذلك بملاءمة البيانات على منحنى اللوجستي، ويستخدم الانحدار اللوجستي عدة متغيرات متوقعة و التي يمكن ان تكون رقمية او فئوية علي سبيل المثال، يستخدم الانحدار اللوجستي في تسويق لحساب توقعات الميل المستهلك الى شراء منتج ما او امتناعه عن الشراء ويستخدم الانحدار اللوجستي بشكل واسع في الطب والعلوم الاجتماعي كما يمكن تعريفه بأنه أسلوب احصائي لفحص العالقة بين المتغير التابع النوعي ومتغير واحد او اكثر من المتغيرات المستقلة، أي انه الأسلوب الإحصائي المستخدم لفحص وتوفيق العالقة بين المتغير التابع النوعي ثنائي القيمة

ومتغير واحد أو أكثر من المتغيرات المستقلة أيا كان نوعها، ويسمي هنا بتحليل الانحدار اللوجستي الثنائي Regression Logistic Binary ويعرف كذلك بأنه ذلك النوع من الانحدار المستخدم في التنبؤ بقيم المتغيرات التابعة النوعية أو الفئوية بالاعتماد على مجموعة متغيرات مستقلة مختلطة، كأن يكون قسم منها متغيرات مستمرة أو قياسات، والقسم الآخر يكو ن على شكل متغيرات متقطعة نوعية أو فئوية .

3-2 تحويلات الانحدار اللوجستي:

يصنف النموذج اللوجستي ضمن النماذج اللاخطية التي يمكن تحويلها الي نماذج خطية

وتسمى هذه النماذج بالنماذج الخطية ضمنياً (جوهرياً)Intrinsically Linear Model

ويميل الإحصائيون عادة الى التحويل الخطي لهذه النماذج لازال انحناءات معلماتها وذلك لتأثير هذه الانحناءات السلبي في حالة وجودها على خصائص مقدرات المربعات الصغرى لها

Least Squares estimators (LSE) ومن ثم الاستجابة التي يتم التنبؤ بها حيث تكون هذه المقدرات في الغالب متحيزة ولا تتوزع طبيعيا وتبايناتها لا تكون أصغر ما يمكن مما يجعل نتائج الاختبارات مضللة وهناك عدة إجراءات تحويلية يمكن أن تقدم مساهمات جدية لحل بعض الصعوبات والتحديات التي تواجه تفسير نموذج الانحدار اللوجستي، وفي هذا الجزء من الدراسة سوف يتم عرض أهم المفاهيم التي تسهم في تقديم الحلول وهي على النحو التالي:

- **2-3-1 لوغاريتم معامل الترجيح Log Odds (Logit).**

إن استخدام الدالة اللوغاريتمية يعمل على ضم وضغط القيم العالية في البيانات، وتوسيع فيها، حيث إن هذا الضغط والتوسيع قد يعمل على تصحيح عدة مشكلات وفرد القيم الصغيرة جدافي البيانات، مثل التواء التوزيع، ووجود مشاهدات متطرفة، وعدم تجانس التباين . ان Log odds ويسمى أيضاً Logit, هو اللوغاريتم الطبيعي لمعاملات الترجيح Odds, فهو تركيب خطي من المتغيرات المستقلة ،ويمكن التعبير عن نموذج الانحدار اللوجستي بالعلاقة الخطية مع لوجت الاحتمال، وفقاً المعادلة التالية:

$$logit = log(odds) = \log\left(\frac{\mathrm{p}}{1-\mathrm{p}}\right) = \alpha + \beta_1 X_1 + \cdots + \beta_p X_p \quad (3.2)$$

$$-\infty < logit < \infty$$

أما الوظيفة الرئيسة لدالة اللوجت فهي السماح بتطبيق الانحدار الخطي عند تحليل العالقات للبيانات ذات المتغيرات التابعة الثنائية ، وعليه فان الانحدار اللوجستي يشير الى نماذج الانحدار التي تتضمن اللوجت كمتغير تابع في المعادلة، ويحسب مقدار التغير في لوغاريتم معامل الترجيح للمتغير التابع وليس في المتغير التابع نفسه كما هو الحال في تحليل الانحدار الخطي

- **2-3-2 معامل الترجيح أو المفاضلة Odds**

ان معامل الترجيح Odds هو عبارة عن طريقة للتعبير عن مدى احتمال حدوث شيء ما مقارنة باحتمال عدم حدوثه، وغالبا ما يتم التعبير عنه على شكل نسبة بين العددين . ويمكن التعبير عنه كما في المعادلة التالية :

$$odds = \left(\frac{p}{1-p}\right) = exp[\alpha + \beta_1 X_1 + \cdots + \beta_p X_p] \quad (4.2)$$

$$o < odds < \infty$$

حيث أن :

Odds : هي معامل ترجيح حدوث الشيء وتكون

P : هي احتمال حدوث ذلك الشيء

$1-p$:هي احتمال عدم حدوث ذلك الشيء

ومن الجدير ذكره أن معامل الترجيح Odds قد حل مشكله الحدود العليا الاحتمال P, بحيث أصبح معامل الترجيح يأخذ اي قيمة من الصفر وحتى ما لانهاية بدلا من P التي تأخذ اي قيمة من الصفر وحتى الواحد الصحيح .

4-2 مزايا وعيوب الانحدار اللوجستي

اولاً: مزايا الانحدار اللوجستي:

1. يتعامل مع المتغيرات الوصفية أو الكمية ويتضمن حدودا للتفاعلات واختبارا لدلالة المعاملات.
2. يعطى الباحث فكرة عن مقدار تأثير كل متغير مستقل على متغير الاستجابة الثنائية.
3. يرتب الانحدار اللوجستي تأثير المتغيرات مما يسمح للباحث بالاستنتاج بأن متغيرا ما أقوى من المتغير الأخر.
4. يعتبر أقل حساسية اتجاه الانحرافات عن اعتدالية التوزيع لمتغيرات الدراسة.
5. تقديرات المعالم Estimation Parameter وفق النموذج اللوجستي تعد مقبولة في ظل غياب بعض القيود المفروضة على الانحدار الخطي واللوغاريتمي.

ثانياً: عيوب الانحدار اللوجسي:

يعاب على الانحدار اللوجستي أنه لا يأخذ الزمن الذي يسبق حدوث الحدث حيث يرى هذا النموذج يواجه مشكلة عدم القدرة على التعامل مع البيانات التي تعتمد على التغير في النقطة الزمنية التي تسبق حدوث الحدث، وتحتوي على بيانات اختفاء Data Censored وخاصة في الدراسات المقطعية Sectional,Cross لذلك كان هناك البديل الأكثر تعاملا مع هذه الحالة وهو انحدار كوكس.

(الفصـــل الثالـــث)

(مقارنه بين الشبكات العصبية الاصطناعية والانحدار اللوجستي في التنبؤ)

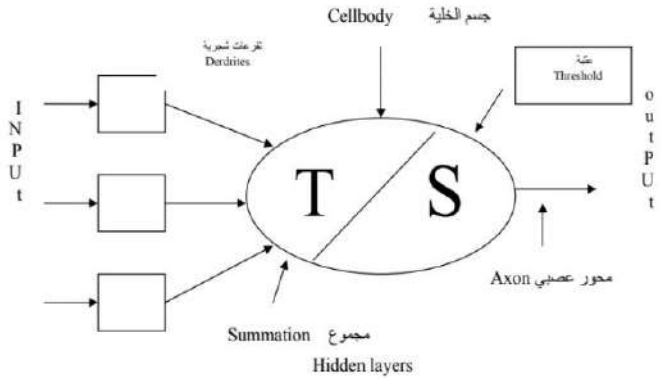
3-1 التنبؤ Prediction

مما لاشك فيه أن عملية تقدير السلوك المستقبلي لأية ظاهرة أو متغير تعد من الأهداف الأساسية للعلوم الإحصائية، وذلك لما لهذه العملية من أهمية كبيرة في عملية التخطيط للظاهرة أو المتغير قيد الدراسة وان التخطيط هو المفتاح لعملية اتخاذ القرار السليم، لقد أولت الحكومات والمؤسسات المختلفة اهتمامات كبيرة لتقدير السلوك المستقبلي للعديد من الظواهر الطبيعية والصناعية والاقتصادية والذي تولد عنه إنشاء المراكز والوحدات المختلفة لهذا الغرض لذا نجد ان موضوع تغير السلوك المستقبلي في استخدام تحليل الانحدار يعد محورا رئيسا اولى الباحثون فيه اهتمامات كبيرة.

3-2 الشبكات العصبية الاصطناعية Artificial Neura Networks

تعد الشبكات العصبية الاصطناعية من التقنيات الذكائية الحديثة التي تشبه في عملها عمل الدماغ البشري والتي تستخدم المحاكاة كطريقة حسابية تشابه إلى حد ما عمل الدماغ في انجاز مهمة معينة وذلك عن طريق معالجة ضخمة موزعة على التوازي ومكونة من وحدات معالجة بسيطة إذ إن هذه الوحدات ما هي إلا عناصر أساسية تسمى (عصبونا) أو عقد (Nodes,Neuron) والتي لها خاصية عصبية إذ أنها تقوم بتخزين المعرفة العملية والمعلومات التجريبية لتجعلها مناعة للمستخدم و ذلك عن طريق ضبط الأوزان، إذ أن ال (Artificial Neural Networks(ANN تتشابه مع الدماغ البشري في أنها تكتسب المعرفة بالتدريب وتخزن هذه المعرفة باستخدام قوى وصل داخل العصبونات تسمى الأوزان وهناك أيضا تشابه عصبي حيوي مما يعطي الفرصة لعلماء البيولوجي في الاعتماد على (ANN) لفهم الظواهر الحيوية. وبعد ذلك تتم معالجة مخرجات كل عقدة بواسطة دالة غير خطية ذات عتبة معينة (Threshold) تعرف بدالة

التنشيط (Acliration function) . والشكل (3-1) يوضح مخطط عمل الشبكة العصبية الاصطناعية.



الشكل (3-1) يوضح مخطط عمل الشبكة العصبية.

3-2-1 النموذج الاحصائي

يعتبر الانحدار من أكثر الأساليب الإحصائية فائدة واستخداما في دراسة العلاقة بين المتغيرات في مختلف المجالات. وهو بهذا المعنى يسمى تحليل الانحدار الخطي Linear Regression Analysis, ويعتبر هذا النموذج الأسهل في التحليل والتنبؤ ،والصورة العامة لنموذج التنبؤ هو:

$$\hat{Y} = F(x) \qquad (1.3)$$

حيث أن (y) متغير تابع (x) متغير مستقل، وينطبق هذا النموذج سواء كانت العلاقة خطية أو غير خطية. ولكي يكون النموذج أكثر تحديدا فان نموذج الخطي يعرف بالمعادلة الآتية:

$$y_i = b_o + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + u_i \quad (2.3)$$

حيث أن (1) المتغير التابع Dependent variable إذن $i = 1,2,3,$ و (x_i) تمثل المتغير المستقل Independent variable إذن $i = 1,2,3$ ويسمى المتغير التفسيري Explanatory لأنه يقدر ما يحدث في المتغير التابع (b_o, b_s) هي معالم النموذج Parameters of Models أو معاملات الانحدار Regression Coefficients وعند تقدير معلمات الانحدار الخطي البسيط وعند تقدير معلمات الانحدار الخطي البسيط $(\hat{b}_o, \hat{b}_s)$ إن الهدف من تحليل الانحدار هو تقدير قيم عددية لمعلمات نموذج الانحدار الخطي $(\hat{b}_o, \hat{b}_s)$ وبتعويض القيم التقديرية للمعلمات $(\hat{b}_o, \hat{b}_s)$ في نموذج الانحدار الخطي الوارد بالمعادلة (2) نحصل على معادلة خط الانحدار التنبؤية Forecasting Equation وعلى الوجه الآتي :

$$\hat{y}_i = \hat{b}_o + \hat{b}_1x_1 + \hat{b}_2x_2 + \hat{b}_3x_3 + \hat{b}_4x_4 + \hat{b}_5x_5 \quad (3.$$

حيث أن :

$\hat{b}_o$:بعد نقطة تقاطع خط الانحدار مع الاحداثي الصادي عن نقطة الأصل .

$\hat{b}_s$:معامل الانحدار أوميل خط الانحدار (Slope) ,s=1,2,3,4,5 .

يعرف معامل الانحدار $\hat{b}_s$ بأنه :

مؤشر إحصائي يقصر مقدار التغير الذي يطرأ على المتغير التابع (yi) ، إذا ما تغير المتغير المستقبل (xi) وحدة واحدة. استخدام نموذج الانحدار المتدرج (Stepwise) . وهنا يتم إدخال المتغيرات المستقلة إلى المعادلة الخطية ، ويتم اختيار متغيرين في الخطوة الأولى لإدخالهما إلى المعادلة ثم تقوم باختيار المتغيرين وإقرار إمكانية استبعادهما وحذفهما في المعادلة ، في كل خطوة تدخل متغيرين آخرين وعليه فإن هذه المعادلة هي أحسن المعادلات التي من الممكن اختيارها. وقد بدأت في الآونة الأخيرة اهتمامات الباحثين في استخدام أساليب الـ شبكات العصبية في المعالجات الإحصائية والتنبؤ باستخدام الشبكات العصبية (By Neural Predictions Network) تستخدم الشبكات العصبية في بعض الأحيان للتنبؤ القصير الأجل خصوصا" في الصناعات الالكترونية وهذه التقنية الحديثة توفر دالة مرنة (flexible function) يمكن أن تلام النماذج غير الخطية (Non–liner), ومنها نماذج الشبكات العصبية النكائية الصناعية والتي تعد دوال غير خطية مرنة والشكل العام لدالة الشبكات العصبية المستخدمة في التنبؤ يتبع الصيغة الاتية:

$$y = F[H[suB][/suB(x)\,, H[suB]2\,[/suB\,](x)\ \ldots\ldots HN(X)] + mu \qquad (4.3)$$

حيث أن :

y : تمثل المتغير التابع dependent variable

X : تمثل المتغيرات التفسيرية أو المتغيرات المستقلة Independent variables

H,F: تمثل دوال الشبكات العصبية.

لما μ فتمثل error trerm حد الخطأ في الدالة.

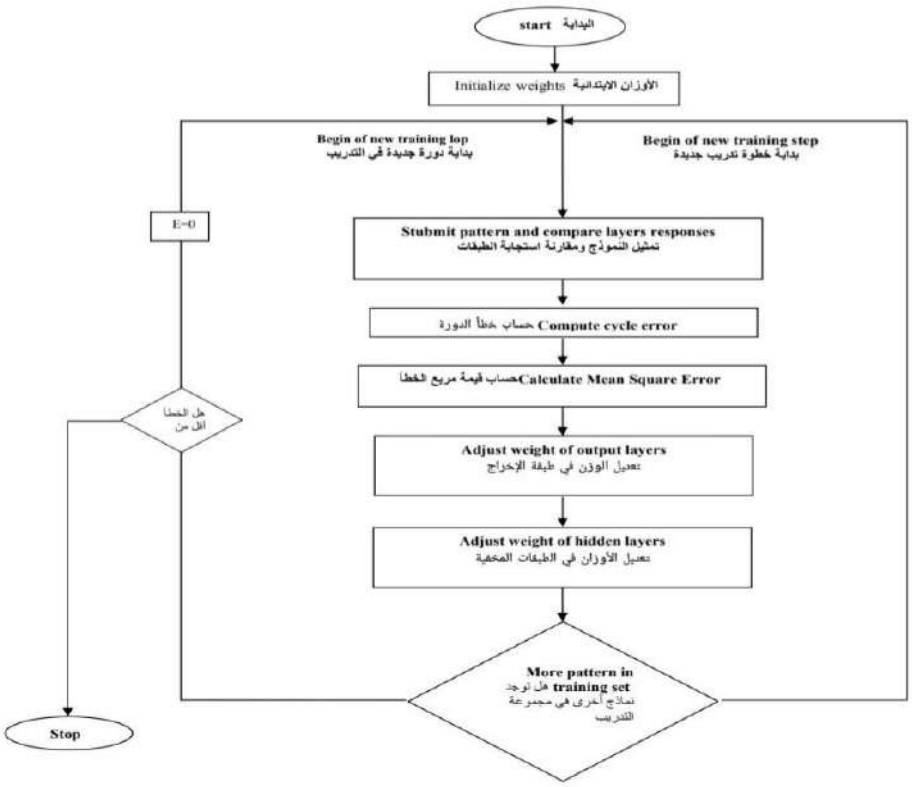
وتسمى مجموعة متغيرات (X) بالمدخلات Input.

ويسمى المتغير (y) بالمخرجات أو الناتج output.

وتسمى H بدوال التحفيز للطبقات الخفية Hidden Layer Activation Function (F) وتمثل مخرجات دالة التحفيز الخفية وهناك شكل محدد specific يستخدم للتنبؤ في هذا النموذج تكون (F) خطية في دوال على شكل حرف S الإنكليزي وتدعى الدالة اللوجستية logistic function وعن نماذج الشبكات العصبية .

1. نموذج شبكات العصبية ذات التغذية الأمامية ذو مخرج واحد single output feed forward neural network .
2. طبقة مخفية واحدة متعددة العقد (one hidden layer with multiple nodes) .
3. دوال التحفيز اللوجستية في الطبقة الخفية (logistic activation function).
4. دوال التحفيز لطبقة المخرج أو الناتج (liner activation function).
5.

يتضح مما سبق وجود علاقة بين نموذج الانحدار ونموذج الشبكات العصبية وأن الفرق بين نموذج الانحدار الخطي ونموذج الشبكات العصبية هو أن دالة الانحدار خطية linear في معلماتها parameters لا توجد هناك دوال للطبقات الخفية في نموذج الانحدار كما في المعادلة (2) سينما مفهوم الشبكات العصبية وحسب المعادلة (4) تحتاج إلى وجدات إدخال الشبكة العصبية اختيار عدد قليل من العقد المخفية عند البدء بتدريب الشبكة العصبية وبعدها تزداد العقد المخفية إلى أن تصل إلى اقل مربع خطا ممكن وأفضل نتائج في معايير المقارنة .ويتم بناء الشبكات العصبية عن طريق البرمجة (أي أن الشبكة العصبونية هي عبارة عن برنامج حاسوبي) يكون من عدد كبير من الدورات التي يزيد استعمالها لحل مشكلة معينة ويوضح الشكل (3-2) خوارزمية تدريب الشبكة ذات الانتشار العكسي.



الشكل (3-2) يوضح خوارزمية تدريب شبكة الانتشار العكسي.

3-3 مفهوم الانحدار اللوجستي

يمكن تعريف نموذج الانحدار اللوجستي بانه نموذج يستخدم للتنبؤ باحتمالية وقوع حدث ما وذلك بملاءمة البيانات على منحنى لوجستي ويستخدم الانحدار اللوجستي عدة متغيرات متوقعة والتي يمكن أن تكون رقمية أو فئوية. على سبيل المثال، يستخدم الانحدار اللوجستي في التسويق لحساب توقعات ميل المستهلك إلى شراء منتج ما أو امتناعه عن الشراء . ويستخدم الانحدار اللوجستي بشكل واسع في الطب والعلوم الاجتماعية.

كما يمكن تعريفه بأنه اسلوب احصائي لفحص العلاقة بين المتغير التابع النوعي ومتغير واحد أو أكثر من المتغيرات المستقلة، أي انه الاسلوب الاحصائي المستخدم لفحص وتوفيق العلاقة بين المتغير التابع النوعي ثنائي القيمة ومتغير واحد أو أكثر من المتغيرات المستقلة ايا كان نوعها ، ويسمى هنا بتحليل الانحدار اللوجستي الثنائي Binary LogisticRegression .ويعرف كذلك بأنه ذلك النوع من الانحدار المستخدم في التنبؤ بقيم المتغيرات التابعة النوعية أو القنوية بالاعتماد على مجموعة متغيرات مستقلة مختلطة ، كان يكون قسم منها متغيرات مستمرة او قياسات، والقسم الآخر يكون على شكل متغيرات متقطعة نوعية أو فئوية. ويقوم نموذج الانحدار اللوجستي على فرض أساسي هو أن المتغير التابع ل الذي تهتم بدراسته هو متغير

ثنائي يأخذ القيمة (1) باحتمال (p) والقيمة (0) باحتمال (1) ، أي إلى حدوث الاستجابة وعدم حدوثها ، وكما تعلم في الانحدار الخطي الذي تأخذ متغيراته المستقلة والمتغير التابع قيماً مستمرة فإن النموذج الذي يربط بين المتغيرات هو كما في المعادلة رقم (5) ادناه

$$y = b_o + b_1 x_1 + e \qquad (5.3)$$

أذان (y) يمثل متغيراً مشاهداً مستمراً وبغرض أن متوسط قيم (y) المشاهدة أو الفعلية عند قيمة معينة للمتغير x هي E (y) وأن المتغير e يمثل الخطأ $e = y - \hat{y}$ فإنه يمكن كتابة النموذج على النحو التالي كما في المعادلة رقم (6) :

$$E\left(\frac{y}{x}\right) = b_o + b_1 x_1 \qquad (6.3)$$

ومن المعروف في الانحدار أن الطرف الأيمن لهذه النماذج يأخذ قيماً من (∞−) الى (+∞) ولكن عندما يكون لدينا متغيران أحدهما وهو المتغير التابع ثنائي ، فإن الانحدار الخطي البسيط لا يكون ملائماً لان $E(y/x) = p(y = 1)$، وبذلك تكون قيمة الطرف الأيمن محصورة ما بين الرقمين (0,1) وبذلك يكون النموذج غير قابل للتطبيق من وجهة نظر الانحدار، وإن إحدى طرق حل هذه المشكلة هو إدخال تحويلة رياضية مناسبة على المتغير التابع (y) ومن المعروف $0 \le p \le 1$ ومن ثم بالنسبة $(\frac{p}{p-1})$ عبارة عن مقدار موجب محصور بين$(0, \infty)$ اي ان $0 \le (\frac{p}{p-1}) \le \infty$ وبأخذ اللوغاريتم الطبيعي لمقدار $(\frac{p}{p-1})$ فإن مجال قيمة يصبح محصورا $-\infty \le ln\ (\frac{p}{p-1}) \le \infty$ وبالتالي يمكن كتابة نموذج الانحدار في حالة متغير مستقل واحد كما في المعادلة رقم(7) أدناه:

$$\left(\frac{p}{p-1}\right)\ln = \mathrm{b_o} + \mathrm{b_1 x} \qquad (7.3)$$

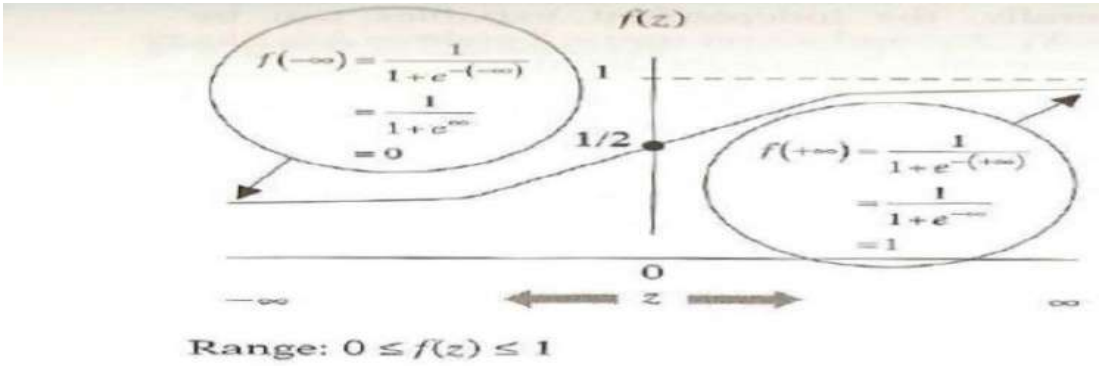
واذا كان لدينا أكثر من متغير مستقل فإن النموذج يأخذ الصيغة كما في المعادلة رقم (8)ادناه:

$$\left(\frac{p}{p-1}\right) ln \sum_{i}^{k} b_j\, x_{ji} \qquad (8.3)$$

ويسمى هذا النموذج بنموذج الانحدار اللوجستي ، وتسمى التحويلة $\ln\left(\frac{p}{p-1}\right)$ بتحويلة لوجت (Logit transformation)، وإن الدالة اللوجستية هي دالة مستمرة تأخذ القيم (100) ، وتقترب (y) من الصفر كلما اقترب الطرف الأيمن للدالة اللوجستية من (∞ −) ، وتقترب (1) من الواحد كلما اقترب الطرف الأيمن لهذه الدالة من (∞) ، وهي دالة متماثلة عندما يكون

الطرف الأيمن لهذه الدالة مساوياً للصفر. ويستخدم الانحدار اللوجستي في التنبؤ باحتمال حدوث حدث معين بتوفيق البيانات بشكل منحنى

لوجستي ، وبالتالي فهو نموذج خطي عام يتخذ شكل الدالة اللوجستية ، كما في الشكل (3-3) ادناه:



شكل (3-3) الدالة اللوجستية حيث يختلف اثر X باختلاف مستوى المتغير Y

ان اعتماد الانحدار في الشكل اعلاه على المنحنى اللوجستي، الذي يأخذ الصيغة كما في المعادلة رقم (9) ادناه:

$$p = \frac{e^{a+bx}}{1+e^{a+bx}} = \frac{1}{1+e^{-(a+bx)}} \qquad (9.3)$$

بدلاً من معادلة الخط المستقيم $(Y = b_o + b_1 x + e)$ هو الذي جعل هذا النوع من الانحدار يسمى بـ (الانحدار اللوجستي).

3-3-1 مبررات استخدام الانحدار اللوجستي

عند محاولة استخدام تحليل الانحدار الخطي بطريقة المربعات الصغرى OLS لتوفيق البيانات ذات المتغيرات التابعة الثنائية ، يبرز نوعان رئيسيان من المشاكل ، هما المبرران لاستخدام الانحدار اللوجستي مكان الانحدار الخطي أو غيره من الاساليب الاحصائية الأخرى لتوفيق البيانات. مع المتغير التابع الثنائي ، وهذان النوعان من المشاكل يمكن وضعهما تحت نوعين رئيسيين هما النوع الأول : مشاكل مفاهيمية والنوع الثاني : مشاكل احصائية . النوع الاول أن ما يرغب الباحث في التنبؤ به في حالة الانحدار اللوجستي ليس هو بالدقة قيمة المتغير التابع ، وانما هو الاحتمال بأن يكون المتغير التابع اما صفرا او واحدا أي اما $\left(p\left(\frac{y}{x}\right)\right) =$

$\left(p\left(\frac{y}{x}\right)\right) = 1$ او 0 وبناء على ذلك ، فان المتغير التابع في حالة الانحدار اللوجستي ليس هو المتغير التابع نفسه كما هو الحال عند استخدام الانحدار الخطي ، وانما هو عبارة عن احتمال ان تكون قيمة المتغير التابع تساوي واحد ، وهو الغالب في الاستخدام، او احتمال ان تكون قيمة المتغير التابع تساوي صفراً . ينشأ النوع الأول من المشاكل الذي يواجه الانحدار الخطي وهو المشاكل المفاهيمية من حقيقة أن الاحتمالات يجب ان تتراوح قيمها بين الواحد الصحيح كحد اعلى والصفر كحد ادنى . أي انه ووفقا لتعريف الاحتمالات لا يمكن لقيمة الاحتمال ان يتجاوز الواح د الصحيح ولا ان ينخفض الى ما دون الصفر . وبما ان تحليل الانحدار الخطي هو نموذج خطي يسمح لخط الانحدار ان يمتد حتى موجب ما لا نهاية او ان يمتد حتى سالب ما لا نهاية، حسب قيمة المتغير او المتغيرات المستقلة ، فإن استخدامه مع البيانات ذات المتغير التابع الثنائي قد يجعل الباحث في حيرة من امره، حيث يجد نفسه امام قيم متوقعة للمتغير التابع تتجاوز الواحد الصحيح او تقل عن الصفر ، الأمر الذي يتناقض تماما مع مفهوم الاحتمالات. وان احد الحلول للمشكلة اعلاه هي اعتماد صيغة القمة والقاع The Floor & Ceiling form ، الموضح في الشكل رقم (١) المذكور انفا ، فوفقا لهذه الصيغة هناك حدود للقيم المتنبأ بها بحيث يفترض الا تتجاوز القيم المتنبأ بها الواحد الصحيح ولا تقل عن الصفر. وبناء على ما جاء اعلاه ، فان توفيق البيانات في حالة المتغير التابع الثنائي لن يكون من خلال استخدام افضل خط مستقيم، ولكن باستخدام المنحنى اللوجستي الذي تقع قيمه بين الصفر والواحد والذي يأخذ الشكل S، وهو الانسب لتوفيق البيانات المشاهدة في حالة المتغيرات التابعة الثنائية. أما النوع الثاني من المشاكل وهو المشاكل الاحصائية ، فيكمن في انتهاك افتراضات تحليل الانحدار الخطي ، واهمها افتراضين رئيسيين هما ، الأول اعتدالية التوزيع الطبيعي Normality والثاني تجانس التباين Homescedaticity ، حيث تنشأ هاتان المشكلتان بسبب الطبيعة الثنائية للمتغير التابع. فبالنسبة لتوزيع الاخطاء سوف لن يكون طبيعيا عند أي مستوى من مستويات X عندما يكون المتغير التابع ثنائيا ، حيث ان توزيع الاخطاء في مثل هذه الحالة سيكون متبعا للتوزيع اللوجستي Logistic Distribution وليس التوزيع الطبيعي Normal Distribution ، كما تنشأ مشكلة عدم تجانس تباين المتغير العشوائي) حد الخطأ (عند استخدام الانحدار الخطي لتوفيق البيانات . المتغير التابع الثنائي ، وذلك لان حد الخطأ في الانحدار يتفاوت ويتغير حسب مستويات المتغير المستقل X ، حيث يلاحظ ان توفيق البيانات من خلال خط مستقيم يمتد من الحد الأدنى للمتغير التابع الى الحد الاعلى سوف يولد قيم اخطاء غير متجانسة.

3-3-2 تقدير وتفسير معاملات الانحدار اللوجستي

من اجل تقدير معاملات الانحدار اللوجستي يتم اللجوء الى طريقة الاحتمال الاعظم Maximum Likelihood Method ، التي تعتبر الطريقة الاكثر ملائمة لكافة النماذج الخطية وغير الخطية ، وتعرف طريقة الاحتمال الاعظم بأنها طريقة تكرارية Iterative تعتمد على تكرار العمليات الحسابية عدة مرات، حتى يتم الوصول الى افضل تقدير للمعاملات، والتي من خلالها يمكن تفسير البيانات المشاهدة . وتستخدم طريقة الاحتمال الاعظم لحساب معاملات اللوجت Logit في الانحدار اللوجستي ، وتهدف هذه الطريقة الى تعظيم لوغاريتم الاحتمال log likelihood ، الذي يعكس مدى امكانية او احتمال ان تكون تلك القيم المشاهدة للمتغير التابع في الامكان توقعها أو التنبؤ بها ، من خلال المتغير او المتغيرات المستقلة ، ويلاحظ ان تقديرات الاحتمال الاعظم انها طريقة تكراري ة تبدأ بقيمة أولية لما ينبغي أن تكون عليه معاملات اللوجت ، ثم تحدد هذه الطريقة اتجاه ومقدار التغير في معاملات اللوجت ، والذي سيزيد من لوغاريتم الاحتمال. واللوجت Logit هو اللوغاريتم الطبيعي لمعاملات الاحتمال Odds ، ويعبر عنه بدلالة الاحتمالات بالصيغة كما في المعادلة رقم (10) ادناه:

$$Logit = \left(\frac{p}{1-p}\right)\ln \qquad (10.3)$$

حيث تسمح لنا دالة اللوجت بتطبيق الانحدار الخطي عند تحميل العالقات للبيانات ذات المتغيرات التابعة الثنائية. اما بالنسبة لتفسير معاملات الانحدار اللوجستي فيتم استخدام معامل اللوجت(logit coefficient) والذي يسمى ايضا بمعامل الانحدار اللوجستي غير المعياري ، ويرمز له بالرمز (b) ، وهو يقابل المعامل غير المعياري (b) في الانحدار الخطي ، ويستخدم المع أمل (b) في الانحدار اللوجستي لتقدير لوغاريتم معامل الترجيح log odds ، بان يكون المتغير التابع يساوي (1) لكل وحدة تغير في المتغير المستقل .علما ان الانحدار اللوجستي يحسب مقدار التغير في لوغاريتم معامل الترجيح log odds للمتغير التابع ، وليس التغير في المتغير التابع نفسه كما هو الأمر في الانحدار الخطي وبالتالي فان تغير معاملات الانحدار اللوجستي بدلالة اللوجت يقدم تفسيرا مطابقا لما هو عليه الأمر في الانحدار الخطي ، ولكن الفرق الوحيد هو في وحدات المتغير التابع ، حيث أن وحدات المتغير التابع في حالة الانحدار اللوجستي تمثل لوغاريتمات معاملات الاحتمال (الترجيح) كما يمكن تفسير معاملات الانحدار اللوجستي بدلالة الاحتمالات ، حيث ان الزيادة في المتغير المستقل بمقدار وحدة واحدة ستزيد اللوجت او لوغاريتم معامل الترجيح log odds او $\ln(\frac{p}{1-p})$ بان يكون المتغير التابع يساوي 1 (Y = 1) بمقدار احتمال معين هو معامل المتغير المستقل.

المصادر

1.ما المقصود بالشبكة العصبونية؟ – شرح "الشبكة العصبونية الاصطناعية" – AWS . https://aws.amazon.com/ar/what-is/neural-network/

2.الشبكات العصبية –مفهومها واستخدامها في التنبؤ –مركز البحوث والدراسات متعدد التخصصات. https://www.mdrscenter.com .

3. كتاب التحليل المتقدم وتنقيب البيانات. د. م. مصطفى عبيد، دار الفكر العربي. القاهرة، الطبيعة الأولى، 2017.

4.الجعضي، خالد (٢٠٠٥) تقنيات صنع القرارات – تطبيقات حاسوبية"، دار الاصحاب للنشر، الرياض، السعودية.

5.عباس، علي (٢٠١٢) ،استخدام نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات الاقتصادية التابعة النوعية " . كلية الإدارة والاقتصاد، جامعة تكريت، العراق، المجلد 2 ،العدد 2.

6. ما المقصود بالانحدار اللوجيستي؟ – شرح نموذج الانحدار اللوجيستي – AWS

. https://aws.amazon.com/ar/what-is/logistic-regression/