Levenberg-Marquardt

د. جورج كرّاز

مقدمة:
إن عملية التدريب بخوارزمية ال Adaline لم تكن الخيار الأفضل في المجال وذلك لأن عامل التعلم لم يكن متغير وإنما ثابت وذلك ما أدى إلى فشله في العديد من النماذج، فلم تكن الطريقة Adaptive أي لم تتلائم مع جميع التطبيقات، هذا ما أدى إلى جعل العلماء يسعون للحصول على Adaptive structure وهي بنية تتلائم مع التطبيقات وتتلائم مع التغييرات.

- بنيت ال Adaptive structure على عدة مبادئ منها تابع الخطأ.
- في الشبكة العصبونية يتم حساب ال Error في البداية ولكل وزن بشكل مفرد، فتم اقتراح فكرة تابع الخطأ Cost function وهو يعبر عن نسبة الخطأ المرتكبة في البنية العصبونية كاملة بشكل عام ويسمى ب mean square error (MSE) وهو تابع يستخدم في الشبكات التي تحل المشاكل الخطية.
- في مشاكل التصنيف تابع ال MSE غير مستخدم وذلك لأن التصنيف يعتمد على قيم اللوغريتمات مثل تابع ال Sigmoid وهو تابع لوغاريتمي بامتياز.
- المشكلة المفروضة علينا هي من تحدد طبيعة التابع الذي سنستخدمه وهو أمر مهم جداً ويؤثر بجودة الشبكة العصبونية.

تمهيد:

بالبدية نعلم أن تابع الخطأ يمثل بالشكل:

$$\boxed{\underset{\theta_0\ \theta_1}{minimize} = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2}$$

حيث:

m: هو عدد العينات

$x^{(i)}$: هي عينة التدريب

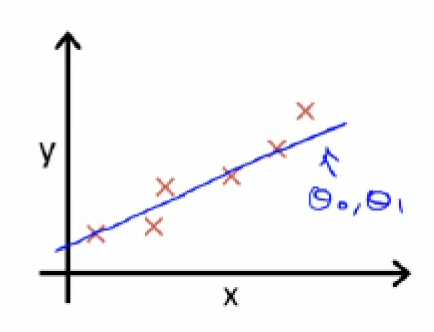
$y^{(i)}$: هي توقع العينة

وحيث:

$$\boxed{h_\theta\left(x^{(i)}\right) = \theta_0 + \theta_1 x}$$

و θ_0 هي ال bias و θ_1 هو الوزن الأول، وبالتالي فهي مجموع الbias مع الوزن جداء عينة التدريب.

ويكتب التابع بالشكل:



$$\boxed{J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2}$$

- ملاحظة: يمكن تعيين الbias بقيمة ثابتة، وذلك في المشاكل الخطية التي يكون التمثيل البياني للعينات على الخط البياني له شكل قريب من المنصف الأول.

وتابع الخطأ هو قطوع ناقصة (دوائر متحدة المركز) والمركز هو ال optimal solution والهدف تصغير هذه الدوائر قدر المستطاع للوصول للمركز، وليس الهدف جعل نسبة الخطأ صفر وإنما قريبة قدر المستطاع من الصفر، وبالتالي يجب تعيين نسبة معينة للتوقف أو تعيين عدد حقب معين للتدريب عليه.

Gradient decent

وهي خوارزمية تساعدني بالوصول لل Local minimum وتعمل على تعديل الأوزان حتى الوصول للهدف وتُمثّل بالعلاقة:

$$repeat\ until\ converengence\ \{$$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0,\theta_1) \quad (for\ j = 0\ and\ \ j = 1)$$

$$\}$$

Learning rate

حيث:

$$\alpha \frac{\partial}{\partial \theta_j} J(\theta_0,\theta_1)$$

تعني المشتق الأول لتابع الخطأ بالنسبة θ

- تذكرة:
- عملية تعديل الأوزان و الbias تتم بشكل موازي:

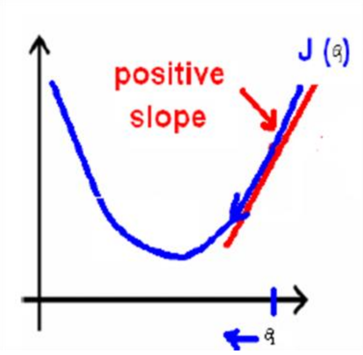
Correct: Simultaneous update

$$temp_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0,\theta_1)$$
$$temp_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0,\theta_1)$$
$$\theta_0 = temp_0$$
$$\theta_1 = temp_1$$

Incorrect:

$$temp_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0,\theta_1)$$
$$\theta_0 = temp_0$$
$$temp_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0,\theta_1)$$
$$\theta_1 = temp_1$$

ولدينا فيها حالتين وهي أن تكون قيمة المشتق أكبر أو أصغر من الصفر.



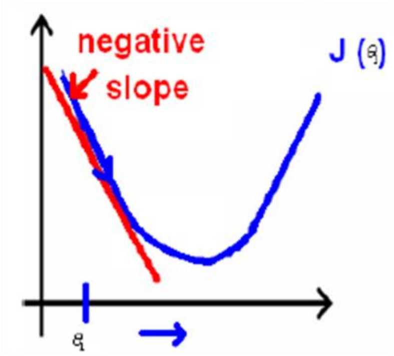
✓ أكبر من الصفر:

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_1)$$

$$\frac{\partial}{\partial \theta_1} J(\theta_1) \geq 0 \Rightarrow \theta_1 := \theta_1 - \alpha(positive)$$

$$\Rightarrow \theta_1 decreases$$

وهذا يعني بأن الوزن الجديد سيكون أقل من القديم



✓ أصغر من الصفر:

$$\frac{\partial}{\partial \theta_1} J(\theta_1) \leq \ 0 \Rightarrow \theta_1 := \theta_1 - \alpha(negative)$$

$$\Rightarrow \theta_1 increases$$

الوزن الجديد سيكون أكبر

في حال لم تكن البنية الثابتة قادرة على حل المشكلة فإننا سنعتمد على بنية Adaptive وتعني بأنها تتغير بتغير أساسات المشكلة وهذا ما اقترحه العالم **Levenberg**.

Levenberg-Marquardt

إن الهدف من تابع الخطأ هو الوصول لأصغر قيمة ممكنة من الخطأ وبشروط محددة، حيث نستخدم خورزميات مثل ال Gradient decent والجاكوبيان وهي خوارزميات تعتمد على المشتقات (الانحدار التدريجي يعتمد على المشتق الثاني والجاكوبيان على المشتق الأول) وذلك للوصول لل Local minimum ولكن في حال لم نتمكن من الوصول سنقوم بالاعتماد على Levenberg-Marquardt method، والفكرة وضع cost function يُأمِّن طريقة للوصول إلى ال local minimum.

ويعتمد على توليد أوتوماتيكي لعامل التعلم Learning rate، وهي من أوسع الخوارزميات القادرة على جعل الحل أمثلي قدر الأمكان.

وتمثل بالعلاقة التالية:

$$E(x,w) = \frac{1}{2}\sum_{p=1}^{P}\sum_{m=1}^{M} e_{p.m}^{2} \quad (1)$$

حيث:

P: يدل على اندكس النمط

M: يدل على اندكس الخرج

i, j: يدلان على اندكس الوزن وهي بين 0 و N و N تدل على عدد الأوزان.

x: مصفوفة الدخل (عينات التدريب)

w: مصفوفة الأوزان

$e_{p.m}$: وهو الخطأ بالخرج عند الخرج m عند تطبيق النمط p ويعطى بالعلاقة:

$$e_{p.m} = d_{p.m} - o_{p.m} \tag{2}$$

حيث d: هي الخرج المرغوب

o: هي الخرج الناتج

Steepest Descent Algorithm

وهو يدخل في معادلة تحديث الأوزان ويعتمد على المشتق الأول ويمكن حسابه بالشكل:

$$g = \frac{\partial E(x,w)}{\partial w} = \begin{bmatrix} \frac{\partial E}{\partial w_1} & \frac{\partial E}{\partial w_2} & \cdots & \frac{\partial E}{\partial w_N} \end{bmatrix}^T \tag{3}$$

لقلب المصفوفة

وتعني المشتق الأول لتابع الخطأ بدلالة x, w وهو يعطينا مصفوفة ولتحويلها لمصفوفة أفقية نقوم بقلبها.

ومعادلة تحديث الأوزان تتم بالعلاقة:

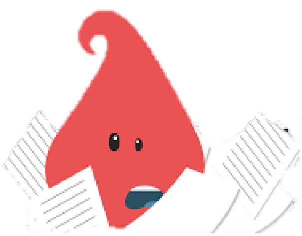
$$w_{k+1} = w_k - \alpha g_k \tag{4}$$

وتعني الوزن الجديد يساوي (الوزن القديم) مطروح منه (قيمة المشتق لتابع الخطأ عند تعديل هذا الوزن) مضروب (بعامل التعلم).

Newton's Method:

تعتمد على المشتق الثاني وتنص على أن جميع الأوزان مرتبطة خطية ويمكن التعبير عنه بالاستفادة من العلاقات السابقة.

لكتابة عناصر المصفوفة المعبرة عن التدرج وبدلالة الw يكون:



$$\begin{cases} g_1 = F_1(w_1, w_2, \ldots, w_N) \\ g_2 = F_2(w_1, w_2, \ldots, w_N) \\ \quad\cdots \\ g_N = F_N(w_1, w_2, \ldots, w_N) \end{cases} \tag{5}$$

حيث كل تدرج عند وزن ما هو تابع w، وكل تدرج يحسب من الذي قبله وبالتالي:

$$\begin{cases} g_1 = g_{1,0} + \frac{\partial g_1}{\partial w_1}\Delta w_1 + \frac{\partial g_1}{\partial w_2}\Delta w_2 + \cdots + \frac{\partial g_1}{\partial w_N}\Delta w_N \\ g_2 = g_{2,0} + \frac{\partial g_2}{\partial w_1}\Delta w_1 + \frac{\partial g_2}{\partial w_2}\Delta w_2 + \cdots + \frac{\partial g_2}{\partial w_N}\Delta w_N \\ \quad\cdots \\ g_N = g_{N,0} + \frac{\partial g_N}{\partial w_1}\Delta w_1 + \frac{\partial g_N}{\partial w_2}\Delta w_2 + \cdots + \frac{\partial g_N}{\partial w_N}\Delta w_N \end{cases} \tag{6}$$

أي من دخل واحد يتفرع لدينا N خرج، وعندها لحساب المشتق لتابع التدرج يكون بالشكل:

$$\frac{\partial g_i}{\partial w_j} = \frac{\partial\left(\frac{\partial E}{\partial w_j}\right)}{\partial w_j} = \frac{\partial^2 E}{\partial w_i \partial w_j} \qquad \boxed{7}$$

الآن بتعويض العلاقة 7 في العلاقة 6:

$$\begin{cases} g_1 = g_{1,0} + \frac{\partial^2 E}{\partial w_1^2}\Delta w_1 + \frac{\partial^2 E}{\partial w_1 \partial w_2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_1 \partial w_N}\Delta w_N \\ g_2 = g_{2,0} + \frac{\partial^2 E}{\partial w_2 \partial w_1}\Delta w_1 + \frac{\partial^2 E}{\partial w_2^2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_2 \partial w_N}\Delta w_N \\ \cdots \\ g_N = g_{N,0} + \frac{\partial^2 E}{\partial w_N \partial w_1}\Delta w_1 + \frac{\partial^2 E}{\partial w_N \partial w_2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_N^2}\Delta w_N \end{cases} \qquad \boxed{8}$$

ولحصول نقطة انعطاف يجب أن يكون $g_1, g_2, g_3 \ldots, g_n$ يساوي الصفر

$$\begin{cases} 0 = g_{1,0} + \frac{\partial^2 E}{\partial w_1^2}\Delta w_1 + \frac{\partial^2 E}{\partial w_1 \partial w_2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_1 \partial w_N}\Delta w_N \\ 0 = g_{2,0} + \frac{\partial^2 E}{\partial w_2 \partial w_1}\Delta w_1 + \frac{\partial^2 E}{\partial w_2^2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_2 \partial w_N}\Delta w_N \\ \cdots \\ 0 = g_{N,0} + \frac{\partial^2 E}{\partial w_N \partial w_1}\Delta w_1 + \frac{\partial^2 E}{\partial w_N \partial w_2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_N^2}\Delta w_N \end{cases} \qquad \boxed{9}$$

بدمج العلاقتين 3 و 9:

$$\begin{cases} -\frac{\partial E}{\partial w_1} = -g_{1,0} + \frac{\partial^2 E}{\partial w_1^2}\Delta w_1 + \frac{\partial^2 E}{\partial w_1 \partial w_2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_1 \partial w_N}\Delta w_N \\ -\frac{\partial E}{\partial w_2} = -g_{2,0} + \frac{\partial^2 E}{\partial w_2 \partial w_1}\Delta w_1 + \frac{\partial^2 E}{\partial w_2^2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_2 \partial w_N}\Delta w_N \\ \cdots \\ -\frac{\partial E}{\partial w_N} = -g_{N,0} + \frac{\partial^2 E}{\partial w_N \partial w_1}\Delta w_1 + \frac{\partial^2 E}{\partial w_N \partial w_2}\Delta w_2 + \cdots + \frac{\partial^2 E}{\partial w_N^2}\Delta w_N \end{cases} \qquad \boxed{10}$$

ولنقم بتمثيلها بمصفوفات لجعلها بشكل أوضح:

$$\begin{bmatrix} -g_1 \\ -g_2 \\ \cdots \\ -g_3 \end{bmatrix} = \begin{bmatrix} -\frac{\partial E}{\partial w_1} \\ -\frac{\partial E}{\partial w_2} \\ \cdots \\ -\frac{\partial E}{\partial w_3} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_1 \partial w_N} \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} & \frac{\partial^2 E}{\partial w_2^2} & \cdots & \frac{\partial^2 E}{\partial w_2 \partial w_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 E}{\partial w_N \partial w_1} & \frac{\partial^2 E}{\partial w_N \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_N^2} \end{bmatrix} \times \begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \cdots \\ \Delta w_N \end{bmatrix} \qquad \boxed{11}$$

وبالتالي حصلنا على تدرج كل وزن بدلالة المشتق الثاني، ولكن إن مصفوفة Hessian تعطى بالشكل:

$$H = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_1 \partial w_N} \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} & \frac{\partial^2 E}{\partial w_2^2} & \cdots & \frac{\partial^2 E}{\partial w_2 \partial w_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 E}{\partial w_N \partial w_1} & \frac{\partial^2 E}{\partial w_N \partial w_2} & \cdots & \frac{\partial^2 E}{\partial w_N^2} \end{bmatrix} \tag{12}$$

وبدمج العلاقة 3 و 12 و 11 ينتج لدينا:

$$-g = H\Delta w \tag{13}$$

أي:

$$\Delta w = -H^{-1} g \tag{14}$$

وبالتالي عملية تحديث الأوزان في طريقة نيوتن تُعطى بالشكل:

$$w_{k+1} = w_k - H_k^{-1} g_k \tag{15}$$

وتعني الوزن الجديد يساوي (الوزن القديم) مطروح منه (مصفوفة هيسيان بالرقم k التي تدل على رقم العينة) مضروباً (بال g_k وهو التدرج عند هذه العينة).

Gauss-Newton Algorithm:

وهي تعتمد على ال Jacobian (أي لها علاقة بالمشتق الأول) وتُعطى المصفوفة بالشكل:

$$J = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial w_1} & \frac{\partial e_{1,1}}{\partial w_2} & \cdots & \frac{\partial e_{1,1}}{\partial w_N} \\ \frac{\partial e_{1,2}}{\partial w_1} & \frac{\partial e_{1,2}}{\partial w_2} & \cdots & \frac{\partial e_{1,2}}{\partial w_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial e_{1,M}}{\partial w_1} & \frac{\partial e_{1,M}}{\partial w_2} & \cdots & \frac{\partial e_{1,M}}{\partial w_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial e_{P,1}}{\partial w_1} & \frac{\partial e_{P,1}}{\partial w_2} & \cdots & \frac{\partial e_{P,1}}{\partial w_N} \\ \frac{\partial e_{P,2}}{\partial w_1} & \frac{\partial e_{P,2}}{\partial w_2} & \cdots & \frac{\partial e_{P,2}}{\partial w_N} \\ \frac{\partial e_{P,3}}{\partial w_1} & \frac{\partial e_{P,3}}{\partial w_2} & \cdots & \frac{\partial e_{P,3}}{\partial w_N} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial e_{P,M}}{\partial w_1} & \frac{\partial e_{P,M}}{\partial w_2} & \cdots & \frac{\partial e_{P,M}}{\partial w_N} \end{bmatrix} \tag{16}$$

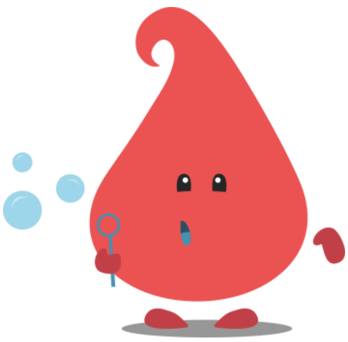
حيث عدد الأعمدة هو عدد العينات، وبجمع المعادلة 1 و 3 فإن التدرج يمكن حسابه بالشكل:

$$g_i = \frac{\partial E}{\partial w_i} = \frac{\left(\frac{1}{2}\sum_{p=1}^{P}\sum_{m=1}^{M} e_{p.m}^2\right)}{\partial w_i} = \sum_{p=1}^{P}\sum_{m=1}^{M}\left(\frac{e_{p.m}}{\partial w_i}\ e_{p.m}\right) \qquad 17$$

وبالتالي من المعادلة 17 أصبح بالإمكان معرفة العلاقة بين مصفوفة الجاكوبيان ومصفوفة التدريج وهي تمثل بالشكل:

$$g = Je \qquad 18$$

حيث:



$$e = \begin{bmatrix} e_{1,1} \\ e_{1,2} \\ \cdots \\ e_{1,M} \\ \cdots \\ e_{P,1} \\ e_{P,2} \\ \cdots \\ e_{P,M} \end{bmatrix} \qquad 19$$

وهي تمثل نسبة الخطأ في كل دورة تعديل أوزان، وبأخذ المعادلة 1 والمعادلة 12 نستنتج أن العنصر بالسطر i وبالعامود j من مصفوفة الهيسيان يحسب بالعلاقة التالية:

$$h_{i,j} = \frac{\partial^2 E}{\partial w_i \partial w_j} = \frac{\partial^2\left(\frac{1}{2}\sum_{p=1}^{P}\sum_{m=1}^{M} e_{p.m}^2\right)}{\partial w_i \partial w_j} = \sum_{p=1}^{P}\sum_{m=1}^{M}\frac{e_{p.m}}{\partial w_i}\ \frac{e_{p.m}}{\partial w_j} + S_{i,j} \qquad 20$$

حيث أن $S_{i,j}$ تعطى بالعلاقة:

$$S_{i,j} = \sum_{p=1}^{P}\sum_{m=1}^{M}\frac{\partial^2 e_{p.m}}{\partial w_i \partial w_j}\ e_{p.m} \qquad 21$$

وهو عدد ثابت قريب من الصفر.

الآن أصبحت العلاقة واضحة بين مصفوفة الهيسيان ومصفوفة الجاكوبيان وهي تُعطى بالشكل:

$$H = J^T J \qquad 22$$

والأن نقوم بدمج العلاقات 15 و 18 و 22 أصبح بإمكاننا التعبير عن علاقة تعديل الأوزان حسب Gauss-Newton بالشكل:

$$w_{k+1} = w_k - (J_k^T J_k)^{-1} J_k e_k \qquad 23$$

وتعني الوزن الجديد يساوي (الوزن القديم) مطروح منه (مقلوب (التحويل الأفقي لمصفوفة جاكوبيان بمصفوفة جاكوبيان)) مضروب (بمصفوفة جاكوبيان(ويعني التدرج) ب قيمة الخطأ).

وتفيد طريقة غاوس-نيوتن بأننا لا نحتاج للمشتقات من الدرجة الثانية لتابع الخطأ عن طريق استخدام مصفوفة جاكوبيان.

ولكن مع ذلك تبقى الخوارزمية تواجه مشكلة في تعقيد مساحة الخطأ، رياضياً يمكن التعبير عن المشكلة أن $j_k^T j$ ربما تكون غير قابلة للقلب.

Levenberg-Marquardt

وهنا جاءت الخوارزمية الخاصة ب Levenberg حيث تهدف لجعل $j_k^T j$ قابلة للقلب، ونعبر عن ذلك بالشكل:

$$H = J^T J + \mu I \qquad (24)$$

حيث:

μ: هو معامل خطي موجب دوماً الهدف منه إعطاء اندكس دائماً

I: مصفوفة القطر الرئيسي لها دوماً له قيمة 1، ويكون عامل التعلم هنا هو مقلوب ال μ.

والآن نقوم بدمج المعادلات 23، 12 و 12, 24 تنتج لدينا قاعد تحديث الأوزان حسب قاعدة ليفينبيرج وتُعطى بالعلاقة:

$$w_{k+1} = w_k - (J_k^T J_k + \mu I)^{-1} J_k e_k \qquad (25)$$

وهذه العلاقة جمعت بين خوارزمية ال Steepest descent و Gauss-Newton وهو اكتشاف ليفينبيرج،

فعندما تكون μ صغير جداً (أي قريب من الصفر) فإن الخوارزمية تستخدم شق خوارزمية Gauss-Newton،

وعندما تكون μ كبيرة جداً شق الخوارزمية Steepest descent يتم استخدامه.

وبالتالي فإن عامل التعلم يعطى بالعلاقة:

$$\alpha = \frac{1}{\mu} \qquad (26)$$

أي مقلوب μ كما تحدثنا سابقاً.

ملخص للقواعد السابقة:

Algorithm	Update Rules	Convergence	Computation Complexity
EBP algorithm	$w_{k+1} = w_k - \alpha g_k$	Stable, slow	Gradient
Newton algorithm	$w_{k+1} = w_k - H_k^{-1} g_k$	Unstable, fast	Gradient and Hessian
Gauss- Newton algorithm	$w_{k+1} = w_k - (J_k^T J_k)^{-1} J_k e_k$	Unstable, fast	Jacobian
Levenberg-Marquardt algorithm	$w_{k+1} = w_k - (J_k^T J_k + \mu I)^{-1} J_k e_k$	Stable, fast	Jacobian
NBN algorithm	$w_{k+1} = w_k - Q_k^{-1} g_k$	Stable, fast	Quasi Hessian

ملاحظة: خوارزمية نيوتن وليفينبيرج هما الأكثر أستخداماً

Pattern association

هي عبارة عن شبكة عصبونية تحوي على طبقة وحيدة وليس بها طبقات مخفية، وعدد ال input وال output متساوي تماماً أي عدد عناصر الدخل يساوي عدد عناصر الخرج.

استخدامها:

تستخدم هذه الشبكة في مقارنة النماذج مع بعضها البعض، حيث يكون لدينا حالتين:

- أن يكون الدخل هو الخرج نفسه تماماً (متطابق) وهذا ما يسمى auto-associative memory.
- أن يكون الدخل مشابه جداً للخرج (قريب) وهذا يسمى hetero-associative memory.

المقصود بال Memory هنا أن المودل سيخزن بالذاكرة مصفوفة التدريب المربعة، في حالة ال hetero -associative memory تكون مصفوفة الأوزان مصفوفة مربعة والقطر الرئيسي لها **يسعى** للواحد (9.2 , 9.9, ...الخ) والمثلث العلوي والمثلث السفلي يجب أن يسعى للصفر أثناء التدريب، أما في حالة auto -associative memory يجب أن يكون العكس (لأننا نرفض النمط المشترك) وبالتالي تكون مصفوفة الأوزان هي مصفوفة مربعة قطرها الرئيسي يسعى للصفر والمثلث العلوي والسفلي يجب أن يسعى ل 1 (وليس قريب جداً من الواحد).

القطر الرئيسي لهذه المصفوفات يسمى cross-talk وتعني الأوزان المشتركة بين الدخل والخرج.
وبالتالي عند تدريب الشبكة الخاصة بال auto-associative نسعى لأن تكون ال cross-talk كبيرة، وعند تدريب شبكة ال hetero-associative نسعى لأن تكون cross-talk صغيرة.
وتكون عملية تحديث الأوزان بالشكل التالي:

$$w_{ij} = \sum_{p=1}^{P} S_i(p)t_j(p)$$

حيث:

S_i: تعني عنصر الدخل

t_j : تعني عنصر الخرج

وبالتالي هي مجموع جداء الدخل بالخرج.

- ملاحظة: تابع الخطأ ليس داخل بعملية التدريب وإنما فقط الدخل والخرج هي المعنية بالعملية.

إلى هنا نكون قد وصلنا إلى نهاية مقرر الشبكات العصبونية، لا تنسونا من صالح دعائكم....