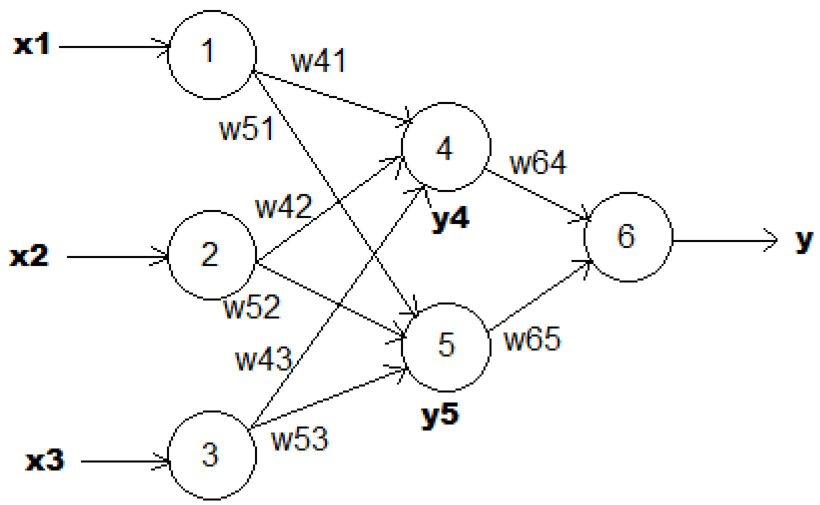
طريقة الانتشار الخلفي للخطأ

Error Back Propagation

بفرض لدينا في الشبكة العصبونية MLP المبينة بالشكل :



جميع العصبونات لها نفس تابع التفعيل **sigmoid**، وتابع الخطأ المستخدم يُعطى بالعلاقة $E=\frac{1}{2}(d-y)^2$

حيث:

d : الخرج المرغوب (خرج المعلم).

y : الخرج الفعلي (خرج الشبكة).

نريد تدريب الشبكة العصبونية باستخدام طريقة الانتشار الخلفي للخطأ Error Back Propagation حيث قُمنا في البداية بإعطاء أوزان الشبكة القيم العشوائية التالية مع إهمال أوزان الانحياز:

$w_{41}=0.2, w_{51}=0.7, \; w_{42}=-0.1, w_{52}=-1.2, \; w_{43}=0.4, w_{53}=1.2, \; w_{64}=1.1, w_{65}=0.1$

ثم قمنا بتغذية الشبكة بعينة التدريب التالية : **[x=(10,30,20) ,d=1]**

المطلوب :

1- اكتب علاقة خرج كل عصبون من عصبونات الشبكة مع حساب قيمته العددية.

2- استنتج مقدار تغيُّر الوزن w_{65} واحسب قيمة هذا الوزن الجديدة ؛علماً أنَّ :

معدل التدريب (Learning Rate) $\eta=0.1$

3- استنتج العلاقة التي تعطي مقدار تغيُّر الوزن w_{42}

الحل :

الطلب الأول :

تتكون طريقة الانتشار الخلفي للخطأ من الخطوات التالية :

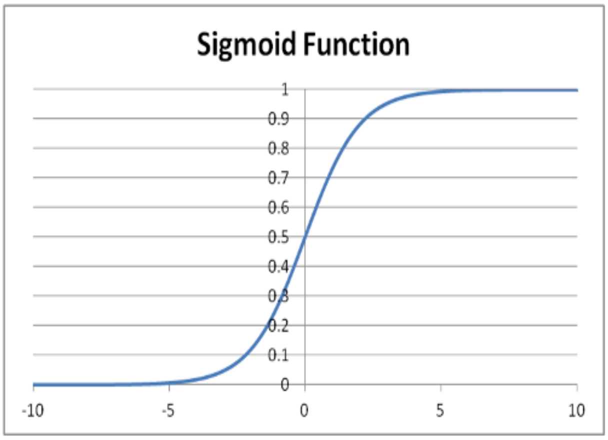
1- إعطاء أوزان عشوائية للوصلات بين عصبونات الشبكة .

2- تغذية الشبكة بإحدى المعطيات المعدّة للتدريب .

3- تطبيق عملية الانتشار الأمامي لتحديد خرج الشبكة العصبونية .

4- مقارنة الخرج الفعلي مع الخرج المطلوب ،وتحديد قيمة الخطأ .

5- التراجع بالخطأ عبر الشبكة وتصحيح الأوزان في الاتجاه الذي يضمن تصغير قيمة الخطأ ومن هنا جاءت تسميتها بالانتشار الخلفي للخطأ .



$$y(x) = \frac{1}{1+e^{-x}}$$

$$y_4 = \frac{1}{1 + e^{-(x_1.w_{41} + x_2.w_{42} + x_3.w_{43})}}$$

$$= \frac{1}{1 + e^{-(10\times0.2 - 30\times0.1 + 20\times0.4)}} = \frac{1}{1 + e^{-7}} = 0.999$$

وبنفس الطريقة :

$$y_5 = \frac{1}{1+\ e^{-(x_1 .w_{51}+x_2 .w_{52}+\ x_3 .w_{53})}}$$

$$= \frac{1}{1+\ \mathrm{e}^{-(10\times0.7\ -\ 30\times1.2\ +\ 20\times1.2)}} = \frac{1}{1+\mathrm{e}^{+5}} = 0.0066$$

$$y = \frac{1}{1+\ \mathrm{e}^{-(\mathrm{y}_4 .\mathrm{w}_{64}\ +\ \mathrm{y}_5 .\mathrm{w}_{65})}}$$

$$= \frac{1}{1+e^{-(0.999\times1.1\ +\ 0.0066\times0.1)}} = \frac{1}{1+\mathrm{e}^{-1.0996}} = 0.750$$

الطلب الثاني :

الآن نقوم بتعديل أوزان طبقة الخرج وبحسب خوارزمية الانحدار التدريجي " gradient descent "

يكون : $$\Delta w_{65} = -\eta \frac{\partial E}{\partial w_{65}}$$

$$\frac{\partial \mathrm{E}}{\partial \mathrm{w}_{65}} = \frac{\partial (\mathrm{d}-\mathrm{y})^2}{\partial \mathrm{y}} \times \frac{\partial \mathrm{y}}{\partial \mathrm{w}_{65}} = -(d-y)\times\frac{\partial y}{\partial w_{65}}$$

أي لإيجاد مشتق تابع الخطأ بالنسبة للأوزان فإنَّنا بحاجة لإيجاد مشتق تابع الخرج y بالنسبة للأوزان ،وبالتالي فإنَّ تابع الخرج يجب أن يكون قابلاً للاشتقاق مثل تابع ال sigmoid.

Chain Rule $\Longrightarrow$ $$\frac{\partial \mathrm{y}}{\partial \mathrm{w}_{65}} = \frac{\partial \mathrm{y}}{\partial \mathrm{s}} \times \frac{\partial \mathrm{s}}{\partial \mathrm{w}_{65}}$$

حيث :

$$s = \sum_{i=4}^{5} w_{6i} \cdot y_i$$

$$\frac{\partial y}{\partial s} = y(1-y)$$

$$\frac{\partial s}{\partial w_{65}} = \frac{\partial}{\partial w_{65}} \sum_{i=4}^{5} w_{6i} \cdot y_i = \frac{\partial}{\partial w_{65}} (w_{64} \cdot y_4 + w_{65} \cdot y_5) = y_5$$

وبالتعويض نجد :

$$\frac{\partial E}{\partial w_{65}} = -(d-y).\, y(1-y).\, y_5$$

حيث : $(d-y).\, y(1-y) = \delta$ ⇐ $\frac{\partial E}{\partial w_{65}} = -\delta.\, y_5$

$$\Delta w_{65} = -\eta \frac{\partial E}{\partial w_{65}} = \eta.\, \delta.\, y_5$$

$$\Delta w_{65} = \eta \,.\, \delta.\, y_5 = 0.1 \times \delta \times 0.0066$$

لكن :

$$\delta = (d-y).\, y(1-y) = (1-0.750) \times (1-0.750) \times 0.750 = 0.0468$$

$$\Delta w_{65} = 3.093 \times 10^{-5}$$

$$w_{65\, new} = w_{65\, old} + \Delta w_{65} = 0.1 + 3.093 \times 10^{-5} = 0.10003$$

نلاحظ أنَّ قيمة الوزن ازدادت بالتالي فإنَّ قيمة الجمع الموزون تزداد ،ومنه فإنَّ الخرج سيقترب من الخرج المطلوب .

الطلب الثالث

$$\Delta w_{42} = -\eta \frac{\partial E}{\partial w_{42}}$$

$$\text{Chain Rule} \Longrightarrow \frac{\partial E}{\partial w_{42}} = \frac{\partial E}{\partial y} \times \frac{\partial y}{\partial w_{42}} = \frac{dE}{dy} \times \frac{dy}{dy_4} \times \frac{\partial y_4}{\partial w_{42}}$$

حيث :

$$\frac{dE}{dy} = \frac{d}{dy}(d-y)^2 = -(d-y)$$

$$\frac{dy}{dy_4} = \frac{\partial}{\partial y_4}\left[\frac{1}{1+e^{-(w_{64}\,.y_4+w_{65}\,.y_5)}}\right] = \frac{w_{64} \times e^{-(w_{64}\,.y_4+w_{65}\,.y_5)}}{(1+e^{-(w_{64}\,.y_4+w_{65}\,.y_5)})^2}$$

$$= w_{64}.\,y(1-y)$$

$$\frac{\partial y_4}{\partial w_{42}} = \frac{\partial}{\partial w_{42}}\left[\frac{1}{1+e^{-(w_{41}\,.x_1+w_{42}\,.x_2+w_{43}\,.x_3)}}\right]$$

$$= \frac{x_2\,.\,e^{-(w_{41}\,.x_1+w_{42}\,.x_2+w_{43}\,.x_3)}}{1+e^{-(w_{41}\,.x_1+w_{42}\,.x_2+w_{43}\,.x_3)}} = x_2\,.\,y_4(1-y_4)$$

ومنه نجد أنَّ :

$$\frac{\partial E}{\partial w_{42}} = -(d-y).\,w_{64}.\,y(1-y).\,x_2.\,y_4(1-y_4)$$

$$\Delta w_{42} = \eta\,.\,x_2.\,y_4(1-y_4).\,w_{64}.\underbrace{y(1-y)(d-y)}_{\delta}$$

$$= \eta\,.\,x_2.\underbrace{y_4(1-y_4).\,w_{64}\,.\,\delta}_{\delta_4}$$

$$\Delta_{42} = \eta\,.\,x_2.\,\delta_4$$