تکلیف شماره ۵ درس شبکه‌های عصبی آرین شکر - ۴۰۰۲۳۴۹۴

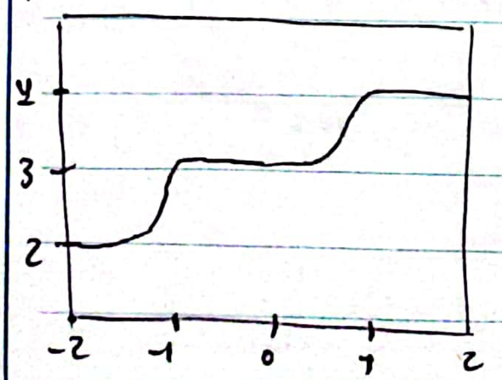
E 11.2. We will use the nominal values presented in the example below figure 11.4, and modify them slightly to fine tune the network response to our desired shape.
$$\left[\begin{array}{l} w^1_{1,1}=10,\ w^1_{2,1}=10,\ b^1_1=-10,\ b^1_2=10, \\ w^2_{1,1}=1,\ w^2_{1,2}=1,\ b^2=0 \end{array}\right]$$

i. The response looks like the nominal response except the entire diagram is displaced by -2 $\Rightarrow$ we set $\underline{b^2=-2}$

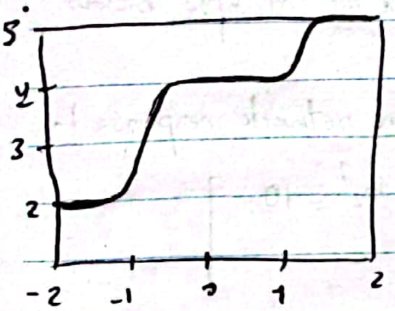
ii. $w^2_{1,2}$ controls the step direction of the left most sigmoid, by setting $\underline{w^2_{1,2}=-1}$ we will flip the first sigmoid while the second one stays the same.

iii. This response looks like (ii) but it is flipped w.r.t. $y=0$. In order to obtain this response this time we will use $w^2_{1,1}$ that controls the step direction of the rightmost sigmoid. If we set $\underline{w^2_{1,1}=-1}$, the second sigmoid is flipped but the first one stays the same.

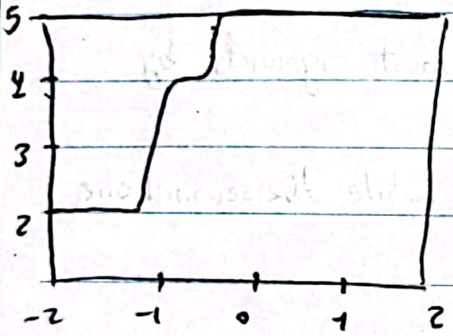
iv. We will use a step by step transformation from the nominal response to get the desired output. First we set $\underline{b^2=2}$ to move the response higher:



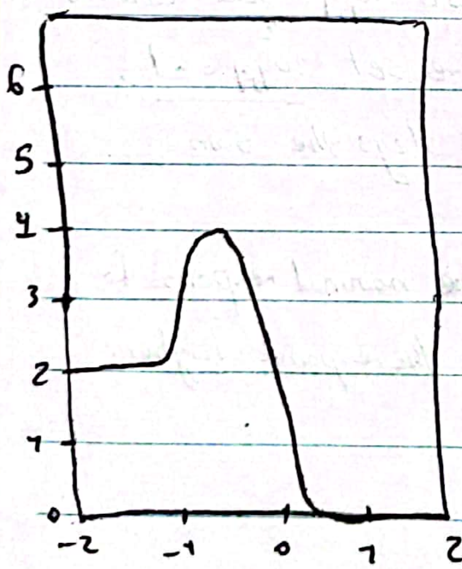
next, we set $w^2_{1,2} = 2$ so that the left sigmoid has a higher peak:



then we bring the step of the next sigmoid close to the first one. specifically we want the step to occur at $-0.5 \rightarrow -b^1_1 / w^1_{1,1} = -0.5 \Rightarrow \boxed{b^1_1 = 5}$



now, all that is left to do is to flip the second sigmoid so that it touches $y=0$. To that end we set $w^2_{1,1} = -4$



$\rightarrow$ which looks like what we want

E11.3. $a^1 = W^1 p + b^1$

$$a^2 = W^2 a^1 + b^2 = W^2(W^1 p) + [W^2 b^1 + b^2]$$

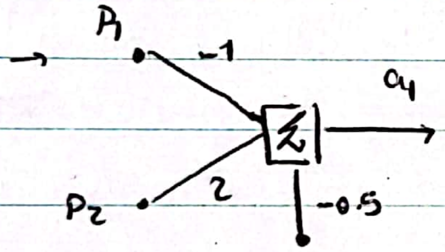
$\rightarrow$ output $= Wp + B$, $W = W^2 W^1$, $B = W^2 b^1 + b^2$

$$\rightarrow W^1 = \begin{bmatrix} -2 & -1 \\ 1 & 3 \end{bmatrix}, \quad W^2 = \begin{bmatrix} 1 & 1 \end{bmatrix} \Rightarrow W = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & -1 \\ 1 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & 2 \end{bmatrix}$$

$$b^1 = \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix}, \quad b^2 = 0.5 \Rightarrow B = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix} + 0.5$$

$$= -0.5$$



$\rightarrow$ This network is equivalent to the original two layer network.

E11.7.

i. $a_1 = W^1 p + b^1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \times 1 + \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$

$a_2 = W^2 a^1 + b^2 = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} + (3) = 0$

$e = t - a = 2 - 0 = 2$

ii. $s^2 = -2\dot{F}^2(n^2)(t-a) = -2[1][2] = -4$

$\dot{f}^2(n^2) = \frac{d}{dn^2}(n^2) = 1$

$$s^1 = \dot{F}^1(n^1)(W^2)^T s^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} [-4] = \begin{bmatrix} 4 \\ -4 \end{bmatrix}$$

$$\dot{f}^1(n^1) = 1 \rightarrow \dot{F}^1(n^1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

iii. $\partial(e)^2 / \partial w^1_{1,1} = s^1_1 \underset{p_1}{a^0_1} = 4 \times 1 = \underline{4}$

E11.10.

Forward pass:

$$n_1 = W^1 p + b^1 = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$a_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \odot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$n_2 = a_2 = W^2 a_1 + b^2 = [2 \;\; 1] \begin{bmatrix} 1 \\ 0 \end{bmatrix} + [-1] = 1$$

$$e = t - a = 2 - 1 = 1$$

Backward Pass:

$$s^2 = -2\dot{F}^2(n^2)(t-a) = -2\left[\dot{f}^2(n^2)\right](t-a) = -2[1][1] = -2$$

$$\dot{f}^2(n^2) = \frac{d}{dn^2}(n^2) = 1$$

$$s^1 = \dot{F}^1(n^1)(W^2)^T s^2 = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} [-2] = \begin{bmatrix} -8 \\ 0 \end{bmatrix}$$

$$\dot{F}^1(n^1) = \begin{bmatrix} \dot{f}^1(n^1_1) & 0 \\ 0 & \dot{f}^1(n^1_2) \end{bmatrix} = \begin{bmatrix} 2n^1_1 & 0 \\ 0 & 2n^1_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

Update: $W^2(1) = W^2(0) - 0.5\, s^2 (a^1)^T$

$$= [2 \;\; 1] - 0.5(-2)[1 \;\; 0] = [3 \;\; 1]$$

$$b^2(1) = b^2(0) - 0.5 s^2 = [-1] - 0.5[-2] = [0]$$

$$w^1(1) = w^1(0) - 0.5\, s^1 (a^0)^T$$

$$= \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} - 0.5 \begin{bmatrix} -8 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 1 & 0 \end{bmatrix}$$

$$b^1(1) = b^1(0) - 0.5 s^1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - 0.5 \begin{bmatrix} -8 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

E 11.13. $a^0 = p$, $a^{m+1} = f^{m+1}\left(\beta^{m+1}\left[w^{m+1} a^m + b^{m+1}\right]\right)$, $m = 0, 1, \ldots, M-1$

$a = a^M$.

Sensitivities are derivatives w.r.t. net input and is therefore invariant w.r.t. changes in the formulation of the net inputs. (only the performance Index affects sensitivities)

$$\Rightarrow s^M = -2\dot{F}^M(n^M)(t-a), \quad s^m = \dot{F}^m(n^m)(w^{m+1})^T s^{m+1}, \quad m = 0, 1, \ldots, M-1$$

However the update rules will change because net inputs have changed.

we have: $n^m = \beta^m \left[w^m a^{m-1} + b^m\right]$

$$\Rightarrow \frac{\partial n^m}{\partial w^m} = \beta^m a^{m-1}, \quad \frac{\partial n^m}{\partial b^m} = \beta^m, \quad \frac{\partial n^m}{\partial \beta^m} = \left[w^m a^{m-1} + b^m\right]$$

$$\Rightarrow \begin{cases} w^m(k+1) = w^m(k) - \alpha s^m (\beta^m a^{m-1})^T \\ b^m(k+1) = b^m(k) - \alpha s^m (\beta^m) \\ \beta^m(k+1) = \beta^m(k) - \alpha s^m (w^m a^{m-1} + b^m)^T \end{cases}$$

E11.16. For the same reason as the previous problem, the equations for sensitivities do not change because $F(x)$ (the performance index) has stayed the same.

Moreover the gradients w.r.t w^1, w^2 and b^1, b^2 are unchanged. The only additional term w.r.t. which we need to backpropagate is $w^{2,1}$:

$$\frac{\partial \hat{F}}{\partial w_{1,j}^{2,1}} = \underbrace{\frac{\partial \hat{F}}{\partial n_i^2}}_{s_i^2} \times \frac{\partial n_i^2}{\partial w_{1,j}^{2,1}} = s_i^2 \times \frac{\partial n_i^2}{\partial w_{i,j}^{2,1}}\,.$$

$$n_i^2 = n_1^2 = n^2 = w_{1,1}^{2,1} \cdot p + w_{1,1}^2 a_1 + b^2 \rightarrow \frac{\partial n^2}{\partial w_{1,1}^{2,1}} = p$$

$$\Rightarrow \boxed{w_{1,1}^{2,1}(k+1) = w_{1,1}^{2,1}(k) - \alpha s^m (p)^T}$$

E11.23. We need the gradient $\frac{\partial}{\partial y(0)} F(y(0)) = \frac{\partial}{\partial y(0)} e^2(K)$.

we define $q(k) = \partial / \partial y(k)\; e^2(K)$, then we would need to calculate $q(0)$

as we have: $\frac{\partial}{\partial y(0)} F(y(0)) = \partial / \partial y(0)\; e^2(K) = q(0)$.

we can derive a recurrence relation for $q(k)$ as follows:

$$q(k-1) = \partial / \partial y(k-1)\; e^2(K) = \frac{\partial e^2(K)}{\partial y(k)} \times \frac{\partial y(k)}{\partial y(k-1)} = q(k) \cdot \frac{\partial (f(y(k-1)))}{\partial y(k-1)}$$

$= q(k) \cdot \dot{f}(y(k-1))$. Also at $k=K$ we have, $q(K) = \frac{\partial e^2(K)}{\partial y(K)}$

$$= \partial / \partial y(K)\; (t - y(K))^2 = -2y(k)(t - y(k))^2.$$

$$\rightarrow \boxed{\begin{array}{l} q(K) = -2y(K)(t - y(K))^2,\quad q(k-1) = q(k) \cdot \dot{f}(y(k-1)),\ \forall k = K-1, \dots, 1 \\ \longrightarrow \partial F(y(0)) / \partial y(0) = q(0) \end{array}}$$