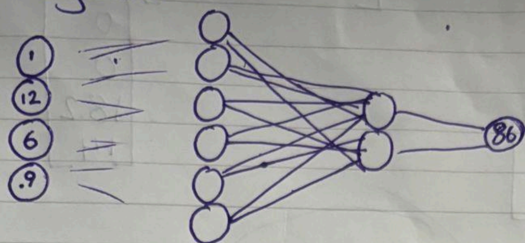
Ankit Adimala
DS 4420
Homework #1

DS 4420 Homework #1 Ankit Adimala

MLP for Regression



1a.) $W^{(1)}$ will be 4x6 and represents the weight of the connection between the 4 input nodes and the 6 nodes of the first hidden layer.

$W^{(2)}$ will be 6x2 and represents the weight of the connection between the 6 nodes of the first hidden layer and the 2 nodes of the second hidden layer.

$W^{(3)}$ will be 2x1 and represent the weight of the connection between the 2 nodes of the second hidden layer and the output node.

1b.) $h^{1pre} = W^{(1)T} x$ $\quad h^{1post} = \mathrm{ReLU}\langle h^{pre} \rangle_+$

$$= \begin{bmatrix} 0 & -1 & 4 & 0 \\ -2 & 2 & -1 & 0 \\ 5 & 0 & -2 & 3 \\ 1 & -3 & -4 & 4 \\ -4 & 1 & 2 & -2 \\ 2 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 12 \\ 6 \\ .9 \end{bmatrix} = \begin{bmatrix} 12 \\ 16 \\ -4.3 \\ -55.4 \\ 18.2 \\ 14 \end{bmatrix} = \begin{bmatrix} 12 \\ 16 \\ 0 \\ 0 \\ 18.2 \\ 14 \end{bmatrix}$$

1c) $h^{2pre} = W^{(2)T} h^{2post} = \begin{bmatrix} -2 & 0 & 1 & -2 & 1 & 0 \\ 1 & -3 & -1 & 6 & 3 & -1 \end{bmatrix} \begin{bmatrix} 12 \\ 16 \\ 0 \\ 0 \\ 18.2 \\ 14 \end{bmatrix} = \begin{bmatrix} -5.8 \\ 4.6 \end{bmatrix}$

$\sigma(h) = \frac{1}{1+e^{-h}}$

$h^{2post} = \begin{bmatrix} \sigma(-5.8) \\ \sigma(4.6) \end{bmatrix}$

$= \begin{bmatrix} \frac{1}{1+e^{5.8}} \\ \frac{1}{1+e^{-4.6}} \end{bmatrix} = \begin{bmatrix} 0.003 \\ 0.990 \end{bmatrix}$

1d) $f_w(x) = W^{(3)T} h^{2post} = \begin{bmatrix} -26 & 96 \end{bmatrix} \begin{bmatrix} 0.003 \\ 0.990 \end{bmatrix} = 94.96$

Error $= 94.96 - 86 = 8.96$

1e.) $\frac{d\mathcal{L}}{dW^{(3)}} = \frac{d\mathcal{L}}{df} \frac{df}{dW^{(3)}}$

$= \frac{1}{n} \sum (f-y)^2 \cdot W^{(3)T} \begin{cases} x & W^{(2)T}x > 0 \\ 0 & \text{O.W.} \end{cases}$

$= \frac{2}{n} \sum (W^{(3)T} \langle W^{(2)T} x \rangle - y) \cdot W^{(3)T} \begin{cases} x & W^{(2)T}x > 0 \\ 0 & \text{O.W.} \end{cases}$

$\cdot 8]$
$\cdot 6]$

1e.) (Cont.) $\frac{d\mathcal{L}}{dW^{(2)}} = \frac{d\mathcal{L}}{df} \frac{df}{dh^{(2)}} \frac{dh^{(2)}}{dW^{(2)}}$

$$\frac{d\mathcal{L}}{df} = \frac{2}{n} \sum \left(W^{(3)T} \langle W^{(2)T} x \rangle - y \right)$$

$f = W^{(3)T} h^{(2)}$ so $\frac{df}{dh^{(2)}} = W^{(3)T}$

$h^{(1)} = \text{ReLU} [W^{(1)T} x \rangle_+$

$h^{(2)} = \sigma (W^{(2)T} h^{(1)})$

$$\frac{dh^{(2)}}{dW^{(2)}} = h^{(1)} \cdot \sigma' (W^{(2)T} h^{(1)})$$

96

$$\frac{d\mathcal{L}}{dW^{(2)}} = \frac{2}{n} \sum \left(W^{(3)T} \langle W^{(2)T} x \rangle - y \right) \cdot W^{(3)T} \cdot h^{(1)} \cdot \sigma' \left(W^{(2)T} h^{(1)} \right)$$