Homework II: I) Polynomial Regression:

(a) $x_1 = 0;$ $x_2 = 1.5;$ $x_3 = 3;$ $x_4 = 4.5;$ $x_5 = 6;$

$\phi_j(x) = x^j$ (3rd degree polynomial regression)

$$\phi([1 \ x_1]) = [1 \ x_1 \ x_1^2 \ x_1^3]$$

$$\Phi = \begin{bmatrix} 1 & 0^1 & 0^2 & 0^3 \\ 1 & 1.5^1 & 1.5^2 & 1.5^3 \\ 1 & 3^1 & 3^2 & 3^3 \\ 1 & 4.5^1 & 4.5^2 & 4.5^3 \\ 1 & 6^1 & 6^2 & 6^3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1.5 & 2.25 & 3.375 \\ 1 & 3 & 9 & 27 \\ 1 & 4.5 & 20.25 & 91.125 \\ 1 & 6 & 36 & 216 \end{bmatrix}$$

(b) $\Phi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1.5 & 2.25 & 3.375 \\ 1 & 3 & 9 & 27 \\ 1 & 4.5 & 20.25 & 91.125 \\ 1 & 6 & 36 & 216 \end{bmatrix}$ $t = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$ $W = \begin{bmatrix} W_0 \\ W_1 \\ W_2 \\ W_3 \end{bmatrix}$

$$y(x, W) = W_0 + W_1 x + W_2 x^2 + W_3 x^3$$

$$E(y(x,W), t) = \sum_{i=1}^{5} \left(y^{(i)} - t^{(i)}\right)^2 \Leftrightarrow E(W, D) = \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right)^2$$

$$\min_W E(W, D) \Rightarrow \frac{\partial E(W,D)}{\partial W} = \begin{bmatrix} \frac{\partial E(W,D)}{\partial W_0} \\ \frac{\partial E(W,D)}{\partial W_1} \\ \frac{\partial E(W,D)}{\partial W_2} \\ \frac{\partial E(W,D)}{\partial W_3} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial W_0} \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right)^2 \\ \frac{\partial}{\partial W_1} \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right)^2 \\ \frac{\partial}{\partial W_2} \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right)^2 \\ \frac{\partial}{\partial W_3} \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right)^2 \end{bmatrix} =$$

$$= \begin{bmatrix} 2 \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right) \Phi_0^{(i)} \\ 2 \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right) \Phi_1^{(i)} \\ 2 \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right) \Phi_2^{(i)} \\ 2 \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right) \Phi_3^{(i)} \end{bmatrix} = 2 \sum_{i=1}^{5} \left(\Phi^{(i)T} W - t^{(i)}\right) \Phi^{(i)} = 2 \Phi^T (\Phi W - t)$$

$$\frac{\partial E(w,D)}{\partial w} = 0 \Leftrightarrow 2\Phi^T(\Phi w - t) = 0 \Leftrightarrow 2\Phi^T\Phi w - 2\Phi^T t = 0 \Leftrightarrow$$

$$\Leftrightarrow 2\Phi^T\Phi w = 2\Phi^T t \Leftrightarrow \Phi^T\Phi w = \Phi^T t \Leftrightarrow w = (\Phi^T\Phi)^{-1}\Phi^T t \Leftrightarrow$$

$$\Leftrightarrow w = \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1{,}5 & 2{,}25 & 3{,}375 \\ 1 & 3 & 9 & 27 \\ 1 & 4{,}5 & 20{,}25 & 91{,}125 \\ 1 & 6 & 36 & 216 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1{,}5 & 2{,}25 & 3{,}375 \\ 1 & 3 & 9 & 27 \\ 1 & 4{,}5 & 20{,}25 & 91{,}125 \\ 1 & 6 & 36 & 216 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1{,}5 & 2{,}25 & 3{,}375 \\ 1 & 3 & 9 & 27 \\ 1 & 4{,}5 & 20{,}25 & 91{,}125 \\ 1 & 6 & 36 & 216 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} =$$

$$= \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1{,}5 & 3 & 4{,}5 & 6 \\ 0 & 2{,}25 & 9 & 20{,}25 & 36 \\ 0 & 3{,}375 & 27 & 91{,}125 & 216 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1{,}5 & 2{,}25 & 3{,}375 \\ 1 & 3 & 9 & 27 \\ 1 & 4{,}5 & 20{,}25 & 91{,}125 \\ 1 & 6 & 36 & 216 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1{,}5 & 3 & 4{,}5 & 6 \\ 0 & 2{,}25 & 9 & 20{,}25 & 36 \\ 0 & 3{,}375 & 27 & 91{,}125 & 216 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 5 & 15 & 67{,}5 & 337{,}5 \\ 15 & 67{,}5 & 337{,}5 & 1792{,}125 \\ 67{,}5 & 337{,}5 & 1792{,}125 & 9871{,}875 \\ 337{,}5 & 1792{,}125 & 9871{,}875 & 55700{,}15625 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1{,}5 & 3 & 4{,}5 & 6 \\ 0 & 2{,}25 & 9 & 20{,}25 & 36 \\ 0 & 3{,}375 & 27 & 91{,}125 & 216 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} 0{,}9857 & -0{,}9921 & 0{,}2857 & -0{,}0247 \\ -0{,}9921 & 2{,}8351 & -1{,}1464 & 0{,}1180 \\ 0{,}2857 & -1{,}1464 & 0{,}5079 & -0{,}0549 \\ -0{,}0247 & 0{,}1180 & -0{,}0549 & 0{,}0061 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1{,}5 & 3 & 4{,}5 & 6 \\ 0 & 2{,}25 & 9 & 20{,}25 & 36 \\ 0 & 3{,}375 & 27 & 91{,}125 & 216 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} 0{,}9857 & 0{,}0571 & -0{,}0857 & 0{,}0571 & -0{,}0143 \\ -0{,}9921 & 1{,}0794 & 0{,}3810 & -0{,}6984 & 0{,}2302 \\ 0{,}2857 & -0{,}4762 & -0{,}0635 & 0{,}4127 & -0{,}1587 \\ -0{,}0247 & 0{,}0494 & 0 & -0{,}0494 & 0{,}0247 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1{,}0571 \\ -1{,}1428 \\ 0{,}1905 \\ 0 \end{bmatrix} \Leftrightarrow w \approx \begin{bmatrix} 1{,}0571 \\ -1{,}1428 \\ 0{,}1905 \\ 0 \end{bmatrix} \Rightarrow \begin{aligned} w_0 &\approx 1{,}0571 \\ w_1 &\approx -1{,}1428 \\ w_2 &\approx 0{,}1905 \\ w_3 &= 0 \end{aligned}$$

1 (c)

$$\frac{\partial E(w)}{\partial w_1} = \sum_{k=1}^{5} (t_k - w_1 \phi_{k1})(\phi_{k1}) + \lambda w_1 =$$

$$= -5,28 + 1,0521\lambda$$

$$\frac{\partial E(w)}{\partial w_2} = \sum_{k=1}^{5} (t_k - w_2 \phi_{k2})(\phi_{k2}) + \lambda w_2 = 56,992 + 1,1528\lambda$$

$$=$$

$$\frac{\partial E(w)}{\partial w_3} = \sum_{k=1}^{5} (t_k - w_3 \phi_{k3})(\phi_{k3}) + \lambda w_3$$

$$= -236,28 + 0,1905\lambda$$

$$\frac{\partial E(w)}{\partial w_4} = \sum_{k=1}^{5} t_k - w_4 \phi_{k4})(\phi_{k4}) + \lambda w_4 =$$

$$= 0$$

$$\frac{\partial E(w)}{\partial w_j^k} = \sum_{k=1}^{5} (t_k - w_j \phi_{kj})(\phi_{kj}) + \lambda w_j$$

(1)(d) $\log(\lambda/2)=0 \Rightarrow \frac{\lambda}{2}=1 \Rightarrow \lambda=2$

$$w_{ridge} = \arg\min_{w} \|y-\Phi w\|_2^2 + \lambda \|w\|_2^2 =$$

$$= (\Phi^T \Phi + \lambda I)^{-1} \Phi^T y =$$

$$= \left[\begin{pmatrix} 5 & 55.5 & 67.5 & 377.502 \\ 55.5 & 2072.25 & 1157.62 & 5582.76 \\ 67.5 & 1157.62 & 1792.12 & 9871.92 \\ 377.502 & 5582.76 & 9871.92 & 55702.5 \end{pmatrix} + \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}\right]^{-1}$$

$$\begin{pmatrix} 1 \\ 7 \\ 27 \\ 189 \end{pmatrix} = \begin{pmatrix} 7 & 55.5 & 67.5 & 377.502 \\ 55.5 & 2074.25 & 1157.62 & 5582.76 \\ 67.5 & 1157.62 & 1794.12 & 9871.92 \\ 377.502 & 5582.76 & 9871.92 & 55702.5 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 7 \\ 27 \\ 189 \end{pmatrix} =$$

$$= \cancel{\left(0.2485436 \quad 0.00368 \quad -0.0065 \right.}$$

$$= \begin{pmatrix} 0.2885 & 0.00368 & -0.0645 & 0.00975 \\ 0.0036 & 0.061565 & -0.0063 & 0.000965 \\ -0.064 & -0.0063 & 0.05998 & -0.0070 \\ 0.009353 & 0.000963 & -0.00961 & 0.00157 \end{pmatrix} \begin{pmatrix} 1 \\ 7 \\ 27 \\ 189 \end{pmatrix} =$$

$$= \begin{bmatrix} 0.72068 \\ 0.0152876 \\ -0.281845 \\ 0.094399 \end{bmatrix}$$

1) a) $\frac{\partial E(w)}{\partial w_j} = \sum_{k=1}^{5} (t_k - \cancel{\phi}\, w_j \phi_{kj})(\phi_{kj}) + \lambda\, \text{sign}(w_j)$

$$\frac{\partial E(w)}{\partial w_1} = \sum_{k=1}^{5} (t_k - w_1 \phi_{k1})(\phi_{k1}) + \lambda\, \text{sign}(w_1)$$

$$= -1{,}28 + \lambda$$

$$\frac{\partial E(w)}{\partial w_2} = 56{,}997 - \lambda$$

$$\frac{\partial E(w)}{\partial w_3} = -2 \cdot 36{,}28 + \lambda$$

$$\frac{\partial E(w)}{\partial w_4} = \lambda$$

(f) LASSO regression lacks a closed form solution because it's non-differentiable.

II) Neural Network NN:

(a) $W^{[1]} = \begin{bmatrix} 1&1&1&1&1 \\ 1&1&1&1&1 \\ 1&1&1&1&1 \end{bmatrix}$ $b^{[1]} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ $W^{[2]} = \begin{bmatrix} 1&1&1 \\ 1&1&1 \end{bmatrix}$ $b^{[2]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$W^{[3]} = \begin{bmatrix} 1&1 \\ 1&1 \end{bmatrix}$ $b^{[3]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ Activation function: $f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} = \tanh(x)$

Error function: Squared error loss: $E(t, x^{[3]}) = \frac{1}{2}\sum_{i=1}^{1}(x^{[3]}-t)^2 =$
$= \frac{1}{2}(x^{[3]}-t)^2$ $\quad \eta = 0.1 \quad X = (1,1,1,1,1)^T \quad t = (1,-1)^T$

Forward propagation: $X^{[0]} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ $\quad z^{[1]} = W^{[1]}X^{[0]} + b^{[1]} =$

$= \begin{bmatrix} 1&1&1&1&1 \\ 1&1&1&1&1 \\ 1&1&1&1&1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix} \Leftrightarrow$

$\Leftrightarrow z^{[1]} = \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix}$

$X^{[1]} = \tanh(z^{[1]}) = \begin{bmatrix} \tanh(5) \\ \tanh(5) \\ \tanh(5) \end{bmatrix} \approx \begin{bmatrix} 0,999909 \\ 0,999909 \\ 0,999909 \end{bmatrix} \Leftrightarrow X^{[1]} \approx \begin{bmatrix} 0,999909 \\ 0,999909 \\ 0,999909 \end{bmatrix}$

$z^{[2]} = W^{[2]}X^{[1]} + b^{[2]} = \begin{bmatrix} 1&1&1 \\ 1&1&1 \end{bmatrix}\begin{bmatrix} 0,999909 \\ 0,999909 \\ 0,999909 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2,999727 \\ 2,999727 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} =$

$= \begin{bmatrix} 2,999727 \\ 2,999727 \end{bmatrix} \Leftrightarrow z^{[2]} = \begin{bmatrix} 2,999727 \\ 2,999727 \end{bmatrix}$ $\quad X^{[2]} = \tanh(z^{[2]}) = \tanh\left(\begin{bmatrix} 2,999727 \\ 2,999727 \end{bmatrix}\right) =$

$= \begin{bmatrix} \tanh(2,999727) \\ \tanh(2,999727) \end{bmatrix} \approx \begin{bmatrix} 0,995052 \\ 0,995052 \end{bmatrix} \Leftrightarrow \boxed{X^{[2]} \approx \begin{bmatrix} 0,995052 \\ 0,995052 \end{bmatrix}}$

$z^{[3]} = W^{[3]}X^{[2]} + b^{[3]} = \begin{bmatrix} 1&1 \\ 1&1 \end{bmatrix}\begin{bmatrix} 0,995052 \\ 0,995052 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1,990104 \\ 1,990104 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} =$

$= \begin{bmatrix} 2,990104 \\ 2,990104 \end{bmatrix} \Leftrightarrow z^{[3]} = \begin{bmatrix} 2,990104 \\ 2,990104 \end{bmatrix}$ $\quad X^{[3]} = \tanh(z^{[3]}) =$

$= \tanh\left(\begin{bmatrix} 2,990104 \\ 2,990104 \end{bmatrix}\right) = \begin{bmatrix} \tanh(2,990104) \\ \tanh(2,990104) \end{bmatrix} \approx \begin{bmatrix} 0,994956 \\ 0,994956 \end{bmatrix} \Leftrightarrow X^{[3]} \approx \begin{bmatrix} 0,994956 \\ 0,994956 \end{bmatrix}$

Backward ~~propagation~~ phase: $\frac{\partial E}{\partial x^{[3]}}(t, x^{[3]}) = \frac{\partial\left(\frac{1}{2}(x^{[3]}-t)^2\right)}{\partial x^{[3]}} =$

$= \frac{\partial\left(\frac{1}{2}(x^{[3]}-t)^2\right)}{\partial (x^{[3]}-t)^2} \frac{\partial (x^{[3]}-t)^2}{\partial (x^{[3]}-t)} \frac{\partial (x^{[3]}-t)}{\partial x^{[3]}} = \left(\frac{1}{2}\right)\left(2(x^{[3]}-t)\right)(1) =$

$= x^{[3]} - t$

$$\frac{\partial x^{[l]}}{\partial z^{[l]}}\left(z^{[l]}\right)=\frac{\partial \tanh\left(z^{[l]}\right)}{\partial z^{[l]}}=\frac{\partial \tanh\left(z^{[l]}\right)}{\partial z^{[l]}}=1-\tanh\left(z^{[l]}\right)^2$$

$$\frac{\partial z^{[l]}}{\partial W^{[l]}}\left(W^{[l]},b^{[l]},x^{[l-1]}\right)=\frac{\partial\left(W^{[l]}x^{[l-1]}+b^{[l]}\right)}{\partial W^{[l]}}=x^{[l-1]}$$

$$\frac{\partial z^{[l]}}{\partial b^{[l]}}\left(W^{[l]},b^{[l]},x^{[l-1]}\right)=\frac{\partial\left(W^{[l]}x^{[l-1]}+b^{[l]}\right)}{\partial b^{[l]}}=1$$

demonstrated in page 11 of Lecture4_Backpropagation.pdf

$$\frac{\partial z^{[l]}}{\partial x^{[l-1]}}\left(W^{[l]},b^{[l]},x^{[l-1]}\right)=\frac{\partial\left(W^{[l]}x^{[l-1]}+b^{[l]}\right)}{\partial x^{[l-1]}}=W^{[l]}$$

$$\delta^{[3]}=\frac{\partial E}{\partial x^{[3]}}\circ\frac{\partial x^{[3]}}{\partial z^{[3]}}=\left(x^{[3]}-t\right)\circ\left(1-\tanh\left(z^{[3]}\right)^2\right)=$$

$$=\left(\begin{bmatrix}0.994956\\0.994956\end{bmatrix}-\begin{bmatrix}1\\-1\end{bmatrix}\right)\circ\left(\begin{bmatrix}1\\1\end{bmatrix}-\begin{bmatrix}\tanh(2.990104)^2\\\tanh(2.990104)^2\end{bmatrix}\right)\approx$$

$$\approx\left(\begin{bmatrix}0.994956\\0.994956\end{bmatrix}-\begin{bmatrix}1\\-1\end{bmatrix}\right)\circ\left(\begin{bmatrix}1\\1\end{bmatrix}-\begin{bmatrix}0.989938\\0.989938\end{bmatrix}\right)\approx\begin{bmatrix}-5.075273\times10^{-5}\\0.020073\end{bmatrix}$$

$$\delta^{[2]}=\left(\frac{\partial z^{[3]}}{\partial x^{[2]}}\right)^T\cdot\delta^{[3]}\circ\frac{\partial x^{[2]}}{\partial z^{[2]}}=\left(W^{[3]}\right)^T\cdot\delta^{[3]}\circ\left(1-\tanh\left(z^{[2]}\right)^2\right)=$$

$$=\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}-5.075273\times10^{-5}\\0.020073\end{bmatrix}\circ\left(\begin{bmatrix}1\\1\end{bmatrix}-\begin{bmatrix}\tanh(2.999727)^2\\\tanh(2.999727)^2\end{bmatrix}\right)\approx$$

$$\approx\begin{bmatrix}1.976396\times10^{-4}\\1.976396\times10^{-4}\end{bmatrix}$$

$$\delta^{[1]}=\left(\frac{\partial z^{[2]}}{\partial x^{[1]}}\right)^T\cdot\delta^{[2]}\circ\frac{\partial x^{[1]}}{\partial z^{[1]}}=\left(W^{[2]}\right)^T\cdot\delta^{[2]}\circ\left(1-\tanh\left(z^{[1]}\right)^2\right)=$$

$$=\begin{bmatrix}1&1\\1&1\\1&1\end{bmatrix}\begin{bmatrix}1.976396\times10^{-4}\\1.976396\times10^{-4}\end{bmatrix}\circ\left(\begin{bmatrix}1\\1\\1\end{bmatrix}-\begin{bmatrix}\tanh(5)^2\\\tanh(5)^2\\\tanh(5)^2\end{bmatrix}\right)\approx\begin{bmatrix}7.194081\times10^{-8}\\7.194081\times10^{-8}\\7.194081\times10^{-8}\end{bmatrix}$$

Perform the updates: $\frac{\partial E}{\partial W^{[1]}}=\delta^{[1]}\cdot\left(\frac{\partial z^{[1]}}{\partial W^{[1]}}\right)^T=\delta^{[1]}\cdot\left(x^{[0]}\right)^T=$

$$=\begin{bmatrix}7.194081\times10^{-8}\\7.194081\times10^{-8}\\7.194081\times10^{-8}\end{bmatrix}\begin{bmatrix}1&1&1&1&1\end{bmatrix}\approx\begin{bmatrix}7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}\\7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}\\7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}\end{bmatrix}$$

$$W^{[1]}=W^{[1]}-\eta\frac{\partial E}{\partial W^{[1]}}=\begin{bmatrix}1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\end{bmatrix}-0.1\begin{bmatrix}7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}\\7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}\\7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}&7.2\times10^{-8}\end{bmatrix}\approx$$

$$\approx\begin{bmatrix}0.999999&0.999999&0.999999&0.999999&0.999999\\0.999999&0.999999&0.999999&0.999999&0.999999\\0.999999&0.999999&0.999999&0.999999&0.999999\end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[1]}} = \delta^{[1]} \cdot \left(\frac{\partial z^{[1]}}{\partial b^{[1]}}\right)^T = \delta^{[1]} = \begin{bmatrix} 7{,}194081 \times 10^{-8} \\ 7{,}194081 \times 10^{-8} \\ 7{,}194081 \times 10^{-8} \end{bmatrix}$$

$$b^{[1]} = b^{[1]} - \eta \frac{\partial E}{\partial b^{[1]}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - 0{,}1 \begin{bmatrix} 7{,}194081 \times 10^{-8} \\ 7{,}194081 \times 10^{-8} \\ 7{,}194081 \times 10^{-8} \end{bmatrix} = \begin{bmatrix} -7{,}194081 \times 10^{-9} \\ -7{,}194081 \times 10^{-9} \\ -7{,}194081 \times 10^{-9} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{[2]}} = \delta^{[2]} \cdot \left(\frac{\partial z^{[2]}}{\partial W^{[2]}}\right)^T = \delta^{[2]} \cdot \left(x^{[1]}\right)^T =$$

$$= \begin{bmatrix} 1{,}976396 \times 10^{-4} \\ 1{,}976396 \times 10^{-4} \end{bmatrix} \begin{bmatrix} 0{,}999909 & 0{,}999904 & 0{,}999909 \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} 0{,}000198 & 0{,}000198 & 0{,}000198 \\ 0{,}000198 & 0{,}000198 & 0{,}000198 \end{bmatrix}$$

$$W^{[2]} = W^{[2]} - \eta \frac{\partial E}{\partial W^{[2]}} \approx \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} 0{,}000198 & 0{,}000198 & 0{,}000198 \\ 0{,}000198 & 0{,}000198 & 0{,}000198 \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} 0{,}999980 & 0{,}999980 & 0{,}999980 \\ 0{,}999980 & 0{,}999980 & 0{,}999980 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[2]}} = \delta^{[2]} \cdot \left(\frac{\partial z^{[2]}}{\partial b^{[2]}}\right)^T = \delta^{[2]} = \begin{bmatrix} 1{,}976396 \times 10^{-4} \\ 1{,}976396 \times 10^{-4} \end{bmatrix}$$

$$b^{[2]} = b^{[2]} - \eta \frac{\partial E}{\partial b^{[2]}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0{,}1 \begin{bmatrix} 1{,}976396 \times 10^{-4} \\ 1{,}976396 \times 10^{-4} \end{bmatrix} = \begin{bmatrix} -1{,}976396 \times 10^{-5} \\ -1{,}976396 \times 10^{-5} \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{[3]}} = \delta^{[3]} \cdot \left(\frac{\partial z^{[3]}}{\partial W^{[3]}}\right)^T = \delta^{[3]} \cdot \left(x^{[2]}\right)^T =$$

$$= \begin{bmatrix} -5{,}075273 \times 10^{-5} \\ 0{,}020073 \end{bmatrix} \begin{bmatrix} 0{,}995052 & 0{,}995052 \end{bmatrix} \approx \begin{bmatrix} -0{,}000050 & -0{,}000050 \\ 0{,}019974 & 0{,}019974 \end{bmatrix}$$

$$W^{[3]} = W^{[3]} - \eta \frac{\partial E}{\partial W^{[3]}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} -0{,}000050 & -0{,}000050 \\ 0{,}019974 & 0{,}019974 \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} 1{,}000005 & 1{,}000005 \\ 0{,}998003 & 0{,}998003 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[3]}} = \delta^{[3]} \cdot \left(\frac{\partial z^{[3]}}{\partial b^{[3]}}\right)^T = \delta^{[3]} = \begin{bmatrix} -5{,}075273 \times 10^{-5} \\ 0{,}020073 \end{bmatrix}$$

$$b^{[3]} = b^{[3]} - \eta \frac{\partial E}{\partial b^{[3]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} -5{,}075273 \times 10^{-5} \\ 0{,}020073 \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} 1{,}000005 \\ 0{,}997993 \end{bmatrix}$$

(b) We can't use the Cross Entropy error because the target values are not [0,1], but [1,-1].

(c) $W^{[1]} = \begin{bmatrix} 1&1&1&1&1&1 \\ 1&1&1&1&1&1 \\ 1&1&1&1&1&1 \end{bmatrix}$ $b^{[1]} = \begin{bmatrix} 0\\0\\0 \end{bmatrix}$ $W^{[2]} = \begin{bmatrix} 1&1&1 \\ 1&1&1 \end{bmatrix}$ $b^{[2]} = \begin{bmatrix} 0\\0 \end{bmatrix}$

$W^{[3]} = \begin{bmatrix} 1&1 \\ 1&1 \end{bmatrix}$ $b^{[3]} = \begin{bmatrix} 1\\1 \end{bmatrix}$

Output units: softmax activation function

Hidden units: hyperbolic tangent activation function (tanh)

Error function: Cross Entropy

$\text{softmax}(z^T) \Rightarrow x_i = \frac{e^{z_i}}{\sum_{k=1}^{d} e^{z_k}}$

$\eta = 0.1$ $x = (1,1,1,1,1)^T$

$t = (1,0)^T$

Forward propagation: $X^{[0]} = \begin{bmatrix} 1\\1\\1\\1\\1\\1 \end{bmatrix}$

$z^{[1]} = W^{[1]} X^{[0]} + b^{[1]} =$

$= \begin{bmatrix} 1&1&1&1&1&1 \\ 1&1&1&1&1&1 \\ 1&1&1&1&1&1 \end{bmatrix} \begin{bmatrix} 1\\1\\1\\1\\1\\1 \end{bmatrix} + \begin{bmatrix} 0\\0\\0 \end{bmatrix} = \begin{bmatrix} 5\\5\\5 \end{bmatrix} \Leftrightarrow$

$\Leftrightarrow z^{[1]} = \begin{bmatrix} 5\\5\\5 \end{bmatrix}$

$X^{[1]} = \tanh(z^{[1]}) = \begin{bmatrix} \tanh(5) \\ \tanh(5) \\ \tanh(5) \end{bmatrix} \approx \begin{bmatrix} 0,999909 \\ 0,999909 \\ 0,999909 \end{bmatrix} \Leftrightarrow X^{[1]} \approx \begin{bmatrix} 0,999909 \\ 0,999909 \\ 0,999909 \end{bmatrix}$

$z^{[2]} = W^{[2]} X^{[1]} + b^{[2]} = \begin{bmatrix} 1&1&1 \\ 1&1&1 \end{bmatrix} \begin{bmatrix} 0,999909 \\ 0,999909 \\ 0,999909 \end{bmatrix} + \begin{bmatrix} 0\\0 \end{bmatrix} = \begin{bmatrix} 2,999727 \\ 2,999727 \end{bmatrix}$

$X^{[2]} = \tanh(z^{[2]}) = \begin{bmatrix} \tanh(2,999727) \\ \tanh(2,999727) \end{bmatrix} \approx \begin{bmatrix} 0,995052 \\ 0,995052 \end{bmatrix} \Leftrightarrow X^{[2]} \approx \begin{bmatrix} 0,995052 \\ 0,995052 \end{bmatrix}$

$z^{[3]} = W^{[3]} X^{[2]} + b^{[3]} = \begin{bmatrix} 1&1 \\ 1&1 \end{bmatrix} \begin{bmatrix} 0,995052 \\ 0,995052 \end{bmatrix} + \begin{bmatrix} 1\\1 \end{bmatrix} \approx \begin{bmatrix} 2,990104 \\ 2,990104 \end{bmatrix}$

$X^{[3]} = \text{softmax}(z^{[3]}) = \begin{bmatrix} \text{softmax}(2,990104) \\ \text{softmax}(2,990104) \end{bmatrix} = \begin{bmatrix} 0,5 \\ 0,5 \end{bmatrix} \Leftrightarrow X^{[3]} = \begin{bmatrix} 0,5 \\ 0,5 \end{bmatrix}$

Backward phase: Cross Entropy: $E(t, X^{[3]}) = -\sum_{i=1}^{d} t_i \log x_i^{[3]}$

$\delta^{[3]} = \frac{\partial E(t, X^{[3]})}{\partial z_i} = \frac{\partial}{\partial z_i}\left(-\sum_{k=1}^{d} t_k \log x_k^{[3]}\right) = -\sum_{k=1}^{d} t_k \frac{\partial}{\partial z_i} \log x_k^{[3]} =$

$= -\sum_{k=1}^{d} t_k \frac{1}{x_k^{[3]}} \frac{\partial}{\partial z_i} x_k^{[3]} = -\sum_{k=i} t_k \frac{1}{x_k^{[3]}} \frac{\partial}{\partial z_i} x_k^{[3]} - \sum_{k \neq i} t_k \frac{1}{x_k^{[3]}} \frac{\partial}{\partial z_i} x_k^{[3]} =$

$= -\sum_{k=i} t_k \frac{x_i^{[3]}(1 - x_i^{[3]})}{x_k^{[3]}} - \sum_{k \neq i} t_k \frac{(-\cancel{x_k^{[3]}} x_i^{[3]})}{\cancel{x_k^{[3]}}} =$

$$= -t_i \frac{\cancel{x_i^{[3]}}\,(1-x_i^{[3]})}{\cancel{x_i^{[3]}}} + \sum_{k\neq i} t_k x_i^{[3]} = -t_i + t_i x_i^{[3]} + \sum_{k\neq i} t_k x_i^{[3]} =$$

$$= -t_i + t_i x_i^{[3]} + x_i^{[3]} \sum_{k\neq i} t_k = -t_i + x_i^{[3]}\left(t_i + \sum_{k\neq i} t_k\right) =$$

$$= -t_i + x_i^{[3]}\left(\sum_{k=1}^{d} t_k\right) = x_i^{[3]} - t_i$$

$$\frac{\partial x^{[l]}}{\partial z^{[l]}}(z^{[l]}) = 1 - \tanh(z^{[l]})^2$$

Demonstrated in page 11 of lecture_4_backpropagation.pdf

$$\frac{\partial z^{[l]}}{\partial W^{[l]}}(W^{[l]}, b^{[l]}, x^{[l-1]}) = x^{[l-1]}$$

$$\frac{\partial z^{[l]}}{\partial b^{[l]}}(W^{[l]}, b^{[l]}, x^{[l-1]}) = 1$$

$$\frac{\partial z^{[l]}}{\partial x^{[l-1]}}(W^{[l]}, b^{[l]}, x^{[l-1]}) = W^{[l]}$$

$$\delta^{[3]} = \begin{bmatrix}\delta_1^{[3]}\\ \delta_2^{[3]}\end{bmatrix} = \begin{bmatrix}x_1^{[3]} - t_1\\ x_2^{[3]} - t_2\end{bmatrix} = \begin{bmatrix}0{,}5-1\\ 0{,}5-0\end{bmatrix} = \begin{bmatrix}-0{,}5\\ 0{,}5\end{bmatrix}$$

$$\delta^{[2]} = \left(\frac{\partial z^{[3]}}{\partial x^{[2]}}\right)^T \cdot \delta^{[3]} \circ \frac{\partial x^{[2]}}{\partial z^{[2]}} = (W^{[3]})^T \cdot \delta^{[3]} \circ \left(1 - \tanh(z^{[2]})^2\right) =$$

$$= \begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}-0{,}5\\0{,}5\end{bmatrix} \circ \left(\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}0{,}990129\\0{,}990129\end{bmatrix}\right) = \begin{bmatrix}0\\0\end{bmatrix} \circ \begin{bmatrix}0{,}009871\\0{,}009871\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

$$\delta^{[1]} = \left(\frac{\partial z^{[2]}}{\partial x^{[1]}}\right)^T \cdot \delta^{[2]} \circ \frac{\partial x^{[1]}}{\partial z^{[1]}} = (W^{[2]})^T \cdot \delta^{[2]} \circ \left(1 - \tanh(z^{[1]})^2\right) =$$

$$= \begin{bmatrix}1&1\\1&1\\1&1\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix} \circ \left(\begin{bmatrix}1\\1\\1\end{bmatrix} - \begin{bmatrix}0{,}999818\\0{,}999818\\0{,}999818\end{bmatrix}\right) = \begin{bmatrix}0\\0\\0\end{bmatrix} \circ \begin{bmatrix}0{,}000182\\0{,}000182\\0{,}000182\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

$$\frac{\partial E}{\partial W^{[1]}} = \delta^{[1]} \cdot \left(\frac{\partial z^{[1]}}{\partial W^{[1]}}\right)^T = \delta^{[1]} \cdot (x^{[0]})^T = \begin{bmatrix}0\\0\\0\end{bmatrix}\begin{bmatrix}1&1&1&1&1\end{bmatrix} = \begin{bmatrix}0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}$$

$$W^{[1]} = W^{[1]} - \eta\left(\frac{\partial E}{\partial W^{[1]}}\right) = \begin{bmatrix}1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\end{bmatrix} - 0{,}1\begin{bmatrix}0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix} = \begin{bmatrix}1&1&1&1&1\\1&1&1&1&1\\1&1&1&1&1\end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[1]}} = \delta^{[1]} \cdot \left(\frac{\partial z^{[1]}}{\partial b^{[1]}}\right)^T = \delta^{[1]} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

$$b^{[1]} = b^{[1]} - \eta\frac{\partial E}{\partial b^{[1]}} = \begin{bmatrix}0\\0\\0\end{bmatrix} - 0{,}1\begin{bmatrix}0\\0\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

$$\frac{\partial E}{\partial W^{[2]}} = \delta^{[2]} \cdot \left(\frac{\partial z^{[2]}}{\partial W^{[2]}}\right)^T = \delta^{[2]} \cdot \left(x^{[1]}\right)^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0{,}999909 & 0{,}999909 & 0{,}999909 \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$W^{[2]} = W^{[2]} - \eta \left(\frac{\partial E}{\partial W^{[2]}}\right) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[2]}} = \delta^{[2]} \cdot \left(\frac{\partial z^{[2]}}{\partial b^{[2]}}\right)^T = \delta^{[2]} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$b^{[2]} = b^{[2]} - \eta \frac{\partial E}{\partial b^{[2]}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0{,}1 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\frac{\partial E}{\partial W^{[3]}} = \delta^{[3]} \cdot \left(\frac{\partial z^{[3]}}{\partial W^{[3]}}\right)^T = \delta^{[3]} \cdot \left(x^{[2]}\right)^T = \begin{bmatrix} -0{,}5 \\ 0{,}5 \end{bmatrix} \begin{bmatrix} 0{,}995052 & 0{,}995052 \end{bmatrix} =$$

$$= \begin{bmatrix} -0{,}497526 & -0{,}497526 \\ 0{,}497526 & 0{,}497526 \end{bmatrix}$$

$$W^{[3]} = W^{[3]} - \eta \frac{\partial E}{\partial W^{[3]}} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} -0{,}497526 & -0{,}497526 \\ 0{,}497526 & 0{,}497526 \end{bmatrix} =$$

$$= \begin{bmatrix} 1{,}049753 & 1{,}049753 \\ 0{,}950247 & 0{,}950247 \end{bmatrix}$$

$$\frac{\partial E}{\partial b^{[3]}} = \delta^{[3]} \cdot \left(\frac{\partial z^{[3]}}{\partial b^{[3]}}\right)^T = \delta^{[3]} = \begin{bmatrix} -0{,}5 \\ 0{,}5 \end{bmatrix}$$

$$b^{[3]} = b^{[3]} - \eta \frac{\partial E}{\partial b^{[3]}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0{,}1 \begin{bmatrix} -0{,}5 \\ 0{,}5 \end{bmatrix} = \begin{bmatrix} 1{,}05 \\ 0{,}95 \end{bmatrix}$$