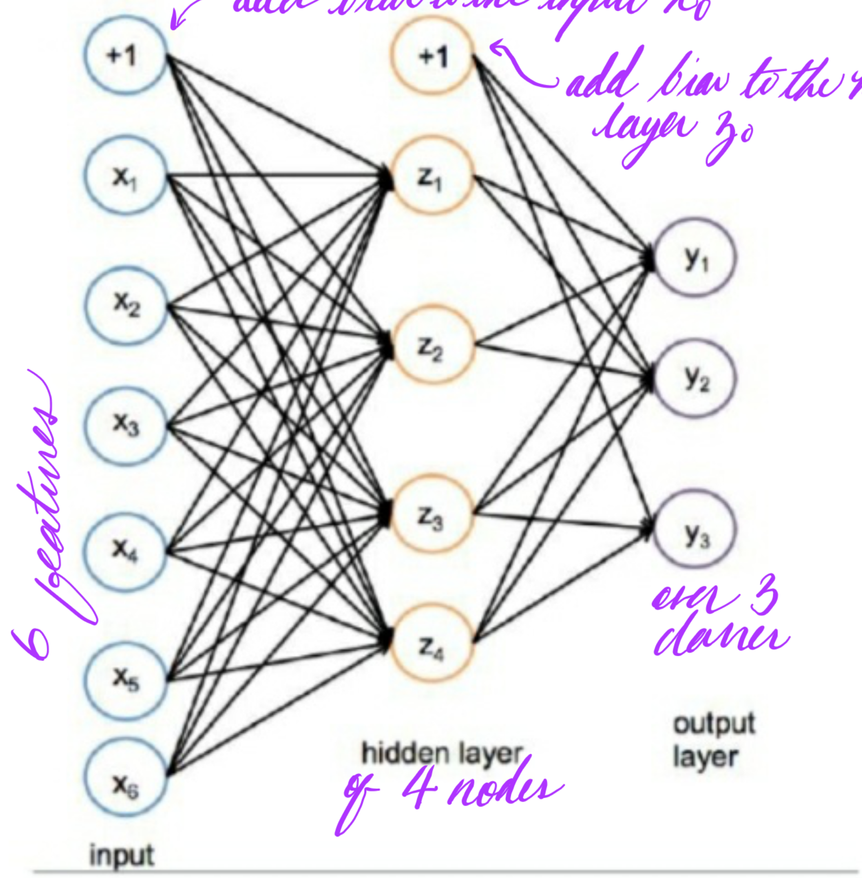
add bias to the input x_0

add bias to the hidden layer z_0 } and then set them to 1



6 features

of 4 nodes

over 3 classes

Fig. 2: A One Hidden Layer Neural Network

α is the matrix of weights from inputs to the hidden layer

β is the matrix of weights from the hidden layer to output layer

α_{ji} weight going to node z_j in the hidden layer from the node x_i in the input layer. ($\alpha_{1,2}$ from x_2 to z_1)

β

Sigmoid activation $\rightarrow$ hidden layer

softmax $\rightarrow$ output layer

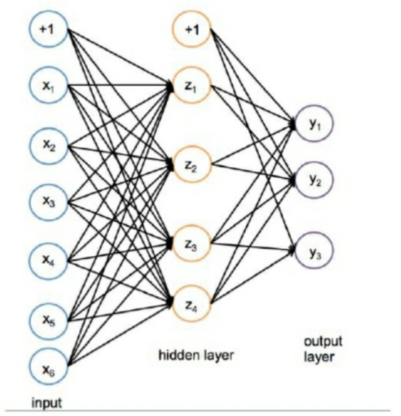


Fig. 2: A One Hidden Layer Neural Network

input: $x = (x_1, x_2, x_3, x_4, x_5, x_6)$

*linear combination at hidden layer: (1)

$$a_j = \alpha_0 + \sum_{i=1}^{6} \alpha_{ji} \times x_i, \forall j \in \{1, 2, 3, 4\}$$

activation at hidden layer: (1)

$$z_j = \sigma(a_j) = \frac{1}{1 + \exp(-a_j)}, \forall j \in \{1, \dots, 4\}$$

*linear combination at output layer: (2)

$$b_K = \beta_0 + \sum_{j=1}^{4} \beta_{K,j} \times z_j, \forall K \in \{1, \dots, 3\}$$

activation at output layer: (2)

$$\hat{y}_K = \frac{\exp(b_K)}{\sum_{l=1}^{3} \exp(b_l)}, \forall K \in \{1, \dots, 3\}$$

*{product of weight_matrix^T with input_vector}

Cross entropy loss, $l(\hat{y}, y)$

output of the network → target output one-hot vector representing the correct class

$$l(\hat{y}, y) = -\sum_{i=1}^{3} y_i \times \log(\hat{y}_i)$$

$\alpha_{1,2}$

when doing prediction, we will predict the argmax of the output layer

$\hat{y}_1 = 0.3$, $\hat{y}_2 = 0.2$, $\hat{y}_3 = 0.5$ ⟶ predict class 3

If the true class from the training data was 2, we would have a one-hot vector y with values

$y_1 = 0$, $y_2 = 1$ $y_3 = 0$

(1,3)

→ We initialize the weights as:

$$\alpha = \begin{bmatrix} 1 & 1 & 2 & -3 & 0 & 1 & -3 \\ 1 & 3 & 1 & 2 & 1 & 0 & 2 \\ 1 & \overset{(2,1)}{2} & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 0 & 2 & 1 & -2 & 2 \end{bmatrix}$$

$x^{(1)} = (1, 1, 0, 0, 1, 1)$

$y^{(1)} = (0, 1, 0)$

run the feed forward of the network over this example (without rounding)

$$\beta = \begin{bmatrix} 1 & 1 & 2 & -2 & 1 \\ 1 & 1 & -1 & 1 & 2 \\ 1 & 3 & 1 & -1 & 1 \end{bmatrix}$$

3×5

$\alpha_{j,0}, \beta_{j,0} = 1$

$x^{(1)} = [1, 1, 1, 0, 0, 1, 1]$

1×7

$$a^{(1)} = \underset{1\times 7}{[1,1,1,0,0,1,1]} \cdot \underset{7\times 4}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 2 & 1 \\ 2 & 1 & 2 & 0 \\ -3 & 2 & 2 & 2 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 2 & -2 \\ -3 & 2 & 1 & 2 \end{bmatrix}} = [2, 7, 8, 2]$$

$$z^{(1)} = \sigma(a_j) = \frac{1}{1+\exp(-a_j)} = [0.88079708, 0.99908895, 0.99966465, 0.88079708]$$

$$z = [1, 0.88079708, 0.99908895, 0.99966465, 0.88079708]$$

$$b^{(1)} = \underset{1\times 5}{z} \cdot \underset{5\times 3}{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 3 \\ 2 & -1 & 1 \\ -2 & 1 & -1 \\ 1 & 2 & 1 \end{bmatrix}} = [2.76044275, 3.64296693, 4.52261261]$$

$$\hat{y}^{(1)} = \frac{\exp(b_k)}{\sum_{l=1}^{3} \exp(b_l)} = [0.10820103, 0.26152113, 0.63027784]$$

$$\ell(\hat{y}, y) = -\sum_{i=1}^{3} y_i \times \log(\hat{y}_i) = -\underset{1\times 3}{[0,1,0]} \cdot \log \underset{3\times 1}{\begin{bmatrix} 0.10820103 \\ 0.26152113 \\ 0.63027784 \end{bmatrix}} = 1.34124021$$

Round to four decimal places for all of problem 3

1a) What is $a_1^{(1)}$ and $z_1^{(1)}$? $a_1^{(1)} = 2$, $z_1^{(1)} \approx 0.8808$

1b) What is $a_3^{(1)}$ and $z_3^{(1)}$? $a_3^{(1)} = 8$, $z_3^{(1)} \approx 0.9997$

1c) What is $b_2^{(1)}$? $b_2^{(1)} \approx 3.6430$

1d) What is $\hat{y_2}^{(1)}$? $\hat{y}_2^{(1)} \approx 0.6303$

1e) Which class would we predict on this example? class 3

1f) What is the total loss on this example? 1.3412

Run backpropagation to update weights

$lr = 1$

$$\ell(\hat{y}, y) = -\sum_{i=1}^{3} y_i \log(\hat{y}_i)$$

$$\frac{\partial \ell}{\partial \hat{y}_i} = -y_i / \hat{y}_i$$

$$\frac{\partial \ell}{\partial \alpha} = \begin{bmatrix} \partial \ell / \partial \alpha_{10} & \cdots & \partial \ell / \partial \alpha_{16} \\ \vdots & \ddots & \vdots \\ \partial \ell / \partial \alpha_{4,0} & \cdots & \partial \ell / \partial \alpha_{46} \end{bmatrix}$$

$$\hat{y}_K = \frac{\exp(b_K)}{\sum_{\ell=1}^{3} \exp(b_\ell)} = \frac{\exp(b_K)}{\exp(b_1) + \exp(b_2) + \exp(b_K)}$$

$$\frac{\partial \ell}{\partial \beta} = \frac{\partial \ell}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b} \cdot \frac{\partial b}{\partial \beta}$$

$$\frac{\partial \hat{y}_K}{\partial b_K} = \frac{(\exp(b_1) + \exp(b_2)) \exp(b_K)}{(\exp(b_K) + \exp(b_1) + \exp(b_2))^2}$$

$$b_K = [z_0, z_1, z_2, z_3, z_4] \cdot \begin{bmatrix} \beta_{1,0} & \beta_{2,0} & \beta_{3,0} \\ \beta_{1,1} & \beta_{2,1} & \beta_{3,1} \\ \beta_{1,2} & \beta_{2,2} & \beta_{3,2} \\ \beta_{1,3} & \beta_{2,3} & \beta_{3,3} \\ \beta_{1,4} & \beta_{2,4} & \beta_{3,4} \end{bmatrix}$$

$$[\underbrace{0.10820103}_{\hat{y}_1}, \underbrace{0.26152113}_{\hat{y}_2}, \underbrace{0.63027784}_{\hat{y}_3}]$$

$$b_1 = z_0 \beta_{1,0} + z_1 \beta_{1,1} + z_2 \beta_{1,2} + z_3 \beta_{1,3} + z_4 \beta_{1,4}$$

$$b_2 = z_0 \beta_{2,0} + z_1 \beta_{2,1} + z_2 \beta_{2,2} + z_3 \beta_{2,3} + z_4 \beta_{2,4}$$

$$b_3 = z_0 \beta_{3,0} + z_1 \beta_{3,1} + z_2 \beta_{3,2} + z_3 \beta_{3,3} + z_4 \beta_{3,4}$$

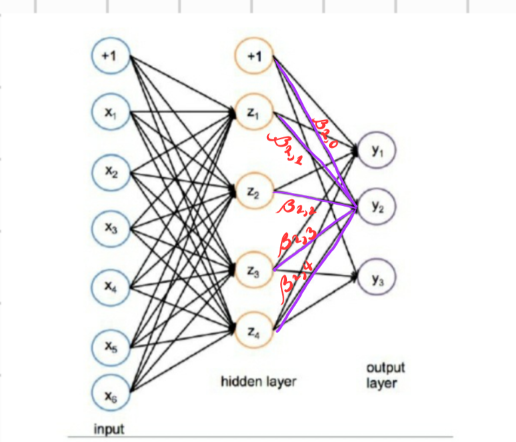


Fig. 2: A One Hidden Layer Neural Network

$$[\underbrace{2.76044275}_{b_1}, \underbrace{3.64296693}_{b_2}, \underbrace{4.52261261}_{b_3}]$$

$$\frac{\partial \ell}{\partial \beta} = \underbrace{\frac{\partial \ell}{\partial \hat{y}}}_{\substack{0 \quad 1 \quad 0 \\ y_1 \quad y_2 \quad y_3}} \cdot \frac{\partial \hat{y}}{\partial b} \cdot \frac{\partial b}{\partial \beta}$$

according to the equation we are calculating the back propagation of class 2

$$\frac{\partial \ell}{\partial \hat{y}_1} = 0 / 0.10820103 = 0 = \frac{\partial \ell}{\partial \hat{y}_3}$$

$$\frac{\partial \ell}{\partial \hat{y}_2} = 1 / 0.26152113 \approx 3.82378286$$

$$\partial \hat{y}_2 / \partial b_2 = \frac{[\exp(b_1) + \exp(b_3)] \exp(b_2)}{[\exp(b_1) + \exp(b_2) + \exp(b_3)]^2} = \frac{[\exp(2.7604) + \exp(4.5226)] \exp(3.6430)}{[\exp(2.7604) + \exp(4.5226) + \exp(3.6430)]^2}$$

$$= 0.19312783$$

$$\frac{\partial \ell}{\partial \beta_{2,0}} = \frac{\partial \ell}{\partial \hat{y}_2} \cdot \frac{\partial \hat{y}_2}{\partial b_2} \cdot \frac{\partial b_2}{\partial \beta_{2,0}}$$

$$\partial b_2 / \partial \beta_{2,0} = z_0, \quad \partial b_2 / \partial \beta_{2,1} = z_1, \quad \frac{\partial b_2}{\partial \beta_{2,2}} = z_2, \quad \partial b_2 / \partial \beta_{2,3} = z_3, \quad \partial b_2 / \partial \beta_{2,4} = z_4$$

$z = [1, 0.88079708, 0.99908895, 0.99966465, 0.88079708]$

$= [z_0, z_1, z_2, z_3, z_4]$

$$\frac{\partial l}{\partial \beta_{2,0}} = 3.82378286 \cdot 0.19312783 \cdot 1 = 0.73847887$$

$$\frac{\partial l}{\partial \beta_{2,1}} = 3.82378286 \cdot 0.19312783 \cdot 0.88079708 = 0.65045003$$

$$\frac{\partial l}{\partial \beta_{2,2}} = 3.82378286 \cdot 0.19312783 \cdot 0.99908895 = 0.73780608$$

$$\frac{\partial l}{\partial \beta_{2,3}} = 3.82378286 \cdot 0.19312783 \cdot 0.99966465 = 0.73823122$$

$$\frac{\partial l}{\partial \beta_{2,4}} = 3.82378286 \cdot 0.19312783 \cdot 0.88079708 = 0.65045003$$

$$\beta = \begin{bmatrix} 1 & 1 & 2 & -2 & 1 \\ 1 & 1 & -1 & 1 & 2 \\ 1 & 3 & 1 & -1 & 1 \end{bmatrix}$$

$\beta_{2,0} := \beta_{2,0} - \eta \, \partial l / \partial \beta_{2,0} = 1 - (1)(0.73847887)$
≈ 0.26152113

$\beta_{2,1} := \beta_{2,1} - \eta \, \partial l / \partial \beta_{2,1} = 1 - (1)(0.65045003)$
≈ 0.34954997

$\beta_{2,1} := \beta_{2,2} - \eta \, \partial l / \partial \beta_{2,2} = 1 - (1)(0.73780608)$
≈ -1.73780608

$\beta_{2,3} := \beta_{2,3} - \eta \, \partial l / \partial \beta_{2,3} = 1 - (1)(0.73823122)$
≈ 0.26176878

$\beta_{2,4} := \beta_{2,4} - \eta \, \partial l / \partial \beta_{2,4} = 1 - (1)(0.65045003)$
≈ 1.34954997

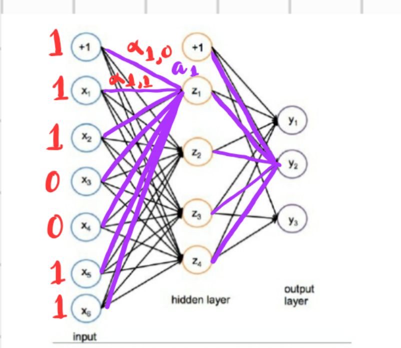


Fig. 2: A One Hidden Layer Neural Network

$$\frac{\partial l}{\partial \alpha} = \frac{\partial l}{\partial z_1} \cdot \frac{\partial z_1}{\partial a_1} \cdot \frac{\partial a_1}{\partial \alpha_{1,0}}$$

a_1, a_2, a_3, a_4

$$l(\hat{y}, y) = -\sum_{i=1}^{3} y_i \log(\hat{y}_i)$$

$$\frac{\partial l}{\partial \hat{y}_i} = -y_i / \hat{y}_i$$

$$\frac{\partial l}{\partial z_1} = \cancelto{0}{\frac{\partial l_1}{\partial z_1}} + \frac{\partial l_2}{\partial z_1} + \cancelto{0}{\frac{\partial l_3}{\partial z_1}}$$

remember it is natural log

$$\frac{\partial l_1}{\partial z_1} = \frac{\partial l_1}{\partial b_1} \cdot \frac{\partial b_1}{\partial z_1}$$

$$l(b_i, y) = -\sum_{i=1}^{3} y_i \frac{b_i}{\sum_{l=1}^{3} b_l}$$

$$l(b_i, y) = -\sum_{i=1}^{3} y_i \log\left(\frac{\exp(b_i)}{\sum_{i=1}^{3} \exp(b_i)}\right)$$

$$\frac{\partial l_2}{\partial b_2} = -1 + \frac{\exp(b_2)}{\sum_{i=1}^{3} \exp(b_i)} = -0.73847887$$

$$\frac{\partial l_2}{\partial z_1} = \frac{\partial l_2}{\partial b_2} \cdot \frac{\partial b_2}{\partial z_1}$$

$$b_2 = z_0\beta_{2,0} + z_1\beta_{2,1} + z_2\beta_{2,2} + z_3\beta_{2,3} + z_4\beta_{2,4}$$

$$\frac{\partial b_2}{\partial z_1} = \beta_{2,1} \qquad \frac{\partial b_2}{\partial z_3} = \beta_{2,3}$$

$$\frac{\partial b_2}{\partial z_2} = \beta_{2,2} \qquad \frac{\partial b_2}{\partial z_4} = \beta_{2,4}$$

$$\frac{\partial l}{\partial z_1} = \frac{\partial l_2}{\partial b_2} \cdot \frac{\partial b_2}{\partial z_1} = -0.25813527$$

$$\frac{\partial l}{\partial z_2} = \frac{\partial l_2}{\partial b_2} \cdot \frac{\partial b_2}{\partial z_2} = 1.28333308$$

$$\frac{\partial l}{\partial z_3} = \frac{\partial l_2}{\partial b_2} \cdot \frac{\partial b_2}{\partial z_3} = -0.19331071$$

$$\frac{\partial l}{\partial z_4} = \frac{\partial l_2}{\partial b_2} \cdot \frac{\partial b_2}{\partial z_4} = -0.99661414$$

$$z_K = \frac{1}{1 + exp(-a_j}$$

$$\frac{\partial z_1}{\partial a_1} = \frac{exp(a_1)}{(exp(a_1) + 1)^2}$$

$$a = [2, 7, 8, 2]$$
$$\quad a_1 \; a_2 \; a_3 \; a_4$$

$$\frac{\partial z_1}{\partial a_1} = 0.10499359$$

$$\frac{\partial z_2}{\partial a_2} = 0.00091022$$

$$\frac{\partial z_3}{\partial a_3} = 0.00033524$$

$$\frac{\partial z_4}{\partial a_4} = 0.10499359$$

$$\frac{\partial a_1}{\partial \alpha_{10}} = \frac{a_1}{\alpha_{1,1}} = \frac{a_1}{\alpha_{1,2}} = \frac{a_1}{\alpha_{1,5}} = \frac{a_1}{\alpha_{1,6}} = 1$$

$$\frac{\partial a_2}{\partial \alpha_{20}} = \frac{a_2}{\alpha_{2,1}} = \frac{a_2}{\alpha_{2,2}} = \frac{a_2}{\alpha_{2,5}} = \frac{a_2}{\alpha_{2,6}} = 1$$

$$\frac{\partial a_3}{\partial \alpha_{30}} = \frac{a_3}{\alpha_{3,1}} = \frac{a_3}{\alpha_{3,2}} = \frac{a_3}{\alpha_{3,5}} = \frac{a_3}{\alpha_{3,6}} = 1$$

$$\frac{\partial a_4}{\partial \alpha_{40}} = \frac{a_4}{\alpha_{4,1}} = \frac{a_4}{\alpha_{4,2}} = \frac{a_4}{\alpha_{4,5}} = \frac{a_4}{\alpha_{4,6}} = 1$$

$$\frac{\partial a_1}{\partial \alpha_{1,3}} = \frac{a_2}{\alpha_{2,3}} = \frac{a_3}{\alpha_{3,3}} = \frac{a_4}{\alpha_{3,4}} = \frac{a_1}{\alpha_{1,4}} = \frac{\partial a_2}{\partial \alpha_{2,4}} = \frac{a_3}{\alpha_{3,4}} = \frac{a_4}{\alpha_{4,4}} = 0$$

$$a_1 = \alpha_{1,0}x_0 + \alpha_{1,1}x_1 + \alpha_{1,2}x_2 + \alpha_{1,3}\cancel{x_3} + \alpha_{1,4}\cancel{x_4} + \alpha_{1,5}x_5 + \alpha_{1,6}x_6$$

$$a_2 = \alpha_{2,0}x_0 + \alpha_{2,1}x_1 + \alpha_{2,2}x_2 + \alpha_{2,3}\cancel{x_3} + \alpha_{2,4}\cancel{x_4} + \alpha_{2,5}x_5 + \alpha_{2,6}x_6$$

$$a_3 = \alpha_{3,0}x_0 + \alpha_{3,1}x_1 + \alpha_{3,2}x_2 + \alpha_{3,3}\cancel{x_3} + \alpha_{3,4}\cancel{x_4} + \alpha_{3,5}x_5 + \alpha_{3,6}x_6$$

$$a_4 = \alpha_{4,0}x_0 + \alpha_{4,1}x_1 + \alpha_{4,2}x_2 + \alpha_{4,3}\cancel{x_3} + \alpha_{4,4}\cancel{x_4} + \alpha_{4,5}x_5 + \alpha_{4,6}x_6$$

$$a_1 = \alpha_{1,0} + \alpha_{1,1} + \alpha_{1,2} + \alpha_{1,5} + \alpha_{1,6}$$

$$a_2 = \alpha_{2,0} + \alpha_{2,1} + \alpha_{2,2} + \alpha_{2,5} + \alpha_{2,6}$$

$$a_3 = \alpha_{3,0} + \alpha_{3,1} + \alpha_{3,2} + \alpha_{3,5} + \alpha_{3,6}$$

$$a_4 = \alpha_{4,0} + \alpha_{4,1} + \alpha_{4,2} + \alpha_{4,5} + \alpha_{4,6}$$

$$\left.\begin{aligned}
\frac{\partial l}{\partial \alpha_{1,0}} &= \frac{\partial l}{\partial z_1} \cdot \frac{\partial z_1}{\partial a_1} \cdot \frac{\partial a_1}{\partial \alpha_{1,0}} \\
\frac{\partial l}{\partial \alpha_{1,1}} &= \frac{\partial l}{\partial z_1} \cdot \frac{\partial z_1}{\partial a_1} \cdot \frac{\partial a_1}{\partial \alpha_{1,1}} \\
\frac{\partial l}{\partial \alpha_{1,2}} &= \frac{\partial l}{\partial z_1} \cdot \frac{\partial z_1}{\partial a_1} \cdot \frac{\partial a_1}{\partial \alpha_{1,2}} \\
\frac{\partial l}{\partial \alpha_{1,3}} &= \frac{\partial l}{\partial z_1} \cdot \frac{\partial z_1}{\partial a_1} \cdot \cancelto{0}{\frac{\partial a_1}{\partial \alpha_{1,3}}} \\
\frac{\partial l}{\partial \alpha_{1,4}} &= \frac{\partial l}{\partial z_1} \cdot \frac{\partial z_1}{\partial a_1} \cdot \cancelto{0}{\frac{\partial a_1}{\partial \alpha_{1,4}}} \\
\frac{\partial l}{\partial \alpha_{1,5}} &= \frac{\partial l}{\partial z_1} \cdot \frac{\partial z_1}{\partial a_1} \cdot \frac{\partial a_1}{\partial \alpha_{1,5}} \\
\frac{\partial l}{\partial \alpha_{1,6}} &= \frac{\partial l}{\partial z_1} \cdot \frac{\partial z_1}{\partial a_1} \cdot \frac{\partial a_1}{\partial \alpha_{1,6}}
\end{aligned}\right\} \text{follows same trend for other}$$

$$\frac{\partial l}{\partial \alpha} = \begin{bmatrix}
\frac{\partial l}{\partial \alpha_{10}} & \frac{\partial l}{\partial \alpha_{11}} & \frac{\partial l}{\partial \alpha_{12}} & \frac{\partial l}{\partial \alpha_{13}} & \frac{\partial l}{\partial \alpha_{14}} & \frac{\partial l}{\partial \alpha_{15}} & \frac{\partial l}{\partial \alpha_{16}} \\
\frac{\partial l}{\partial \alpha_{20}} & \frac{\partial l}{\partial \alpha_{21}} & \frac{\partial l}{\partial \alpha_{22}} & \frac{\partial l}{\partial \alpha_{23}} & \frac{\partial l}{\partial \alpha_{24}} & \frac{\partial l}{\partial \alpha_{25}} & \frac{\partial l}{\partial \alpha_{26}} \\
\frac{\partial l}{\partial \alpha_{30}} & \frac{\partial l}{\partial \alpha_{31}} & \frac{\partial l}{\partial \alpha_{32}} & \frac{\partial l}{\partial \alpha_{33}} & \frac{\partial l}{\partial \alpha_{34}} & \frac{\partial l}{\partial \alpha_{35}} & \frac{\partial l}{\partial \alpha_{36}} \\
\frac{\partial l}{\partial \alpha_{40}} & \frac{\partial l}{\partial \alpha_{41}} & \frac{\partial l}{\partial \alpha_{42}} & \frac{\partial l}{\partial \alpha_{43}} & \frac{\partial l}{\partial \alpha_{44}} & \frac{\partial l}{\partial \alpha_{45}} & \frac{\partial l}{\partial \alpha_{46}}
\end{bmatrix}$$

$$\underbrace{\begin{bmatrix} -0.02710255 & -0.02710255 & -0.02710255 & 0 & 0 & -0.02710255 & -0.02710255 \\ 0.00116812 & 0.00116812 & 0.00116812 & 0 & 0 & 0.00116812 & 0.00116812 \\ -0.00006481 & -0.00006481 & -0.00006481 & 0 & 0 & -0.00006481 & -0.00006481 \\ -0.10463809 & -0.10463809 & -0.10463809 & 0 & 0 & -0.10463809 & -0.10463809 \end{bmatrix}}_{\partial l / \partial \alpha}$$

$$\alpha := \alpha - \eta \, \partial l / \partial \alpha$$

$$\alpha := \begin{bmatrix} 1.02710255 & 1.02710255 & 2.02710255 & 3 & 0 & 1.02710255 & -2.97289745 \\ 0.99883188 & 2.99883188 & 0.99883188 & 2 & 1 & -0.00116812 & 1.99883188 \\ 1.00006481 & 2.00006481 & 2.00006481 & 2 & 2 & 2.00006481 & 1.00006481 \\ 1.10463809 & 1.10463809 & 1.10463809 & 2 & 1 & -1.89536191 & 2.10463809 \end{bmatrix}$$

take same steps to calculate prediction

$$a_j = [x_1, x_2, x_3, x_4, x_5, x_6] \cdot \alpha^T = [2.13551273 \;\; 6.99415942 \;\; 8.00032403 \;\; 2.52319046]$$

$$z_j = \sigma(a_j) = \frac{1}{1 + \exp(-a_j)}, \forall j \in \{1, \dots, 4\} = [0.89430722 \;\; 0.99908362 \;\; 0.99966467 \;\; 0.92575165]$$

$$z = [1 \;\; 0.89430722 \;\; 0.99908362 \;\; 0.99966467 \;\; 0.92575165]$$

$$b = z \cdot \begin{bmatrix} \beta_{1,0} & \beta_{2,0} & \beta_{3,0} \\ \beta_{1,1} & \beta_{2,1} & \beta_{3,1} \\ \beta_{1,2} & \beta_{2,2} & \beta_{3,2} \\ \beta_{1,3} & \beta_{2,3} & \beta_{3,3} \\ \beta_{1,4} & \beta_{2,4} & \beta_{3,4} \end{bmatrix} = [2.81889658 \;\; 0.34894172 \;\; 4.60809216]$$

$$\hat{y}_k = \frac{\exp(b_k)}{\sum_{l=1}^{3} \exp(b_l)}, \forall k \in \{1, \dots, 3\} = [0, 1, 0]$$

Now **use the results of the previous question** to run backpropagation over the network and update the weights. Use learning rate = 1. Do your backpropagation calculations without rounding, and then in your responses, **round to four decimal places**.

$[\beta_{1,0} \; \beta_{2,0} \; \beta_{3,0}] \approx [1, 1.2615, 1]$

2a) What is the updated value of $\beta_{2,1}$? ≈ 0.3495

2b) What is the updated weight of the hidden layer bias term applied to y_1 (e.g. $\beta_{1,0}$)?

2c) What is the updated value of $\alpha_{3,4}$? 2

2d) What is the updated weight of the input layer bias term applied to z_2 (e.g. $\alpha_{2,0}$)?

2e) Once we've updated all of our weights if we ran feed forward over the same example again, which class would we predict? Class 2

$[\alpha_{1,0} \; \alpha_{2,0} \; \alpha_{3,0} \; \alpha_{4,0}] \approx [1.0271 \;\; 0.9988 \;\; 1.0000 \;\; 1.1046]$