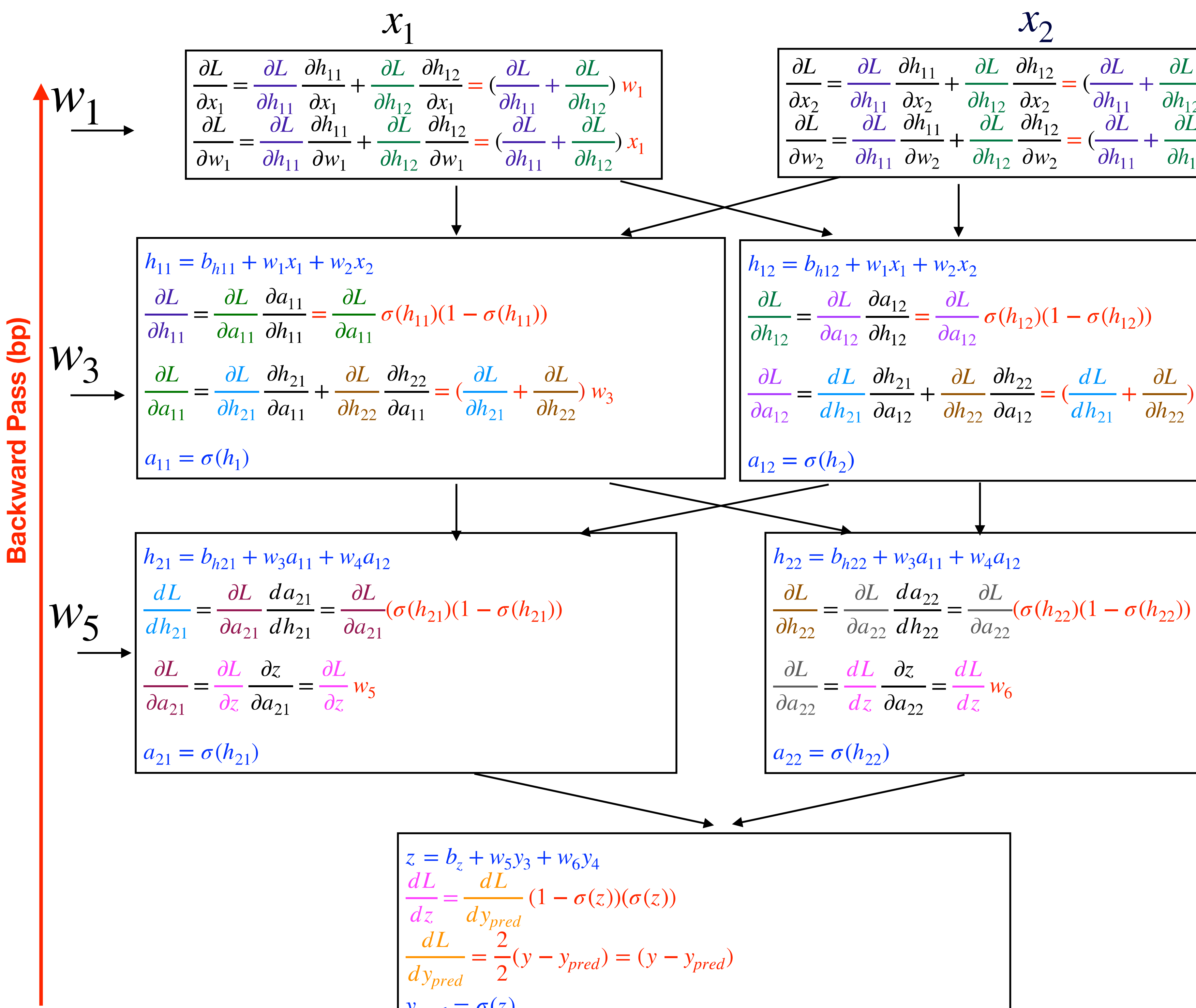
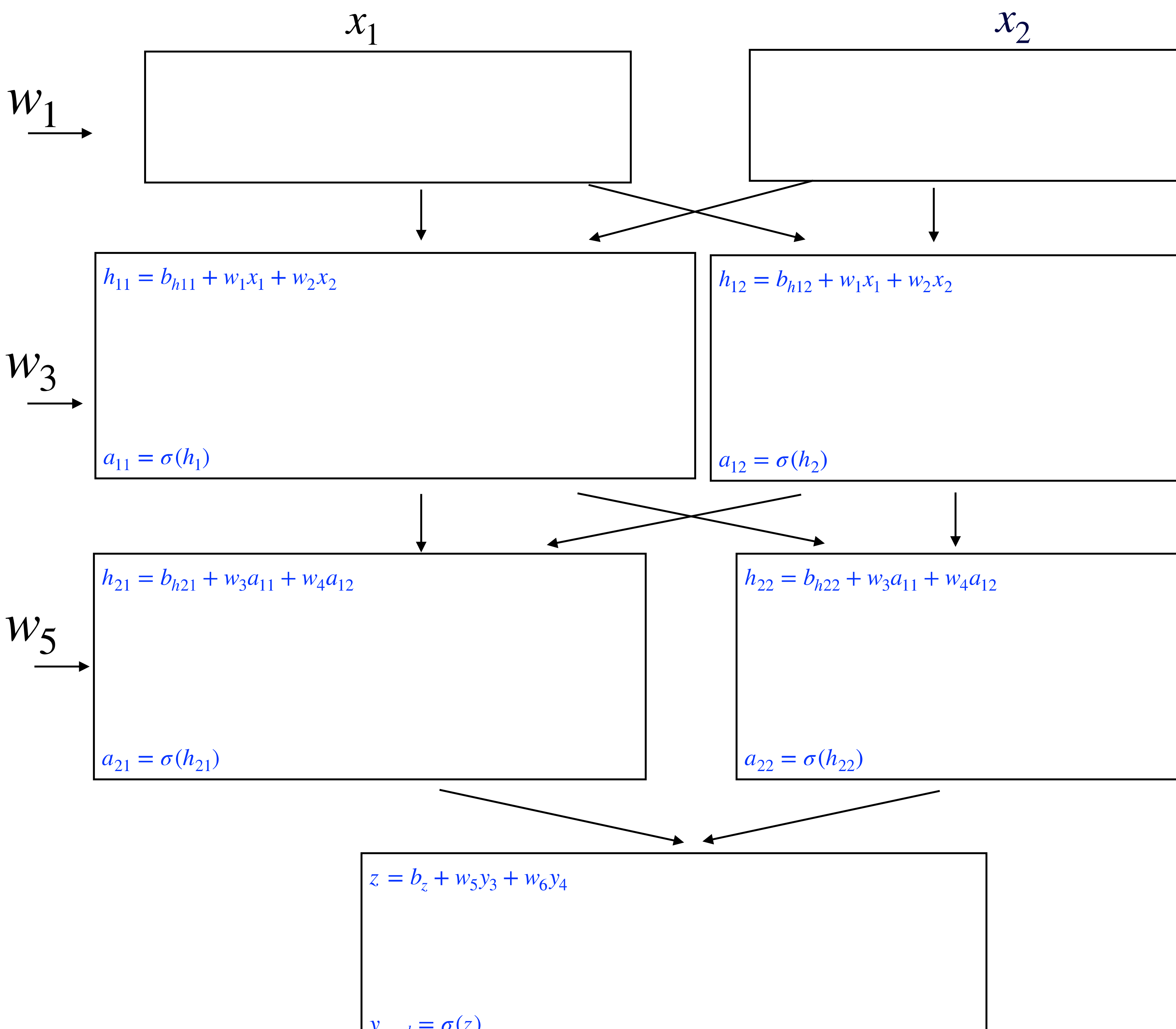
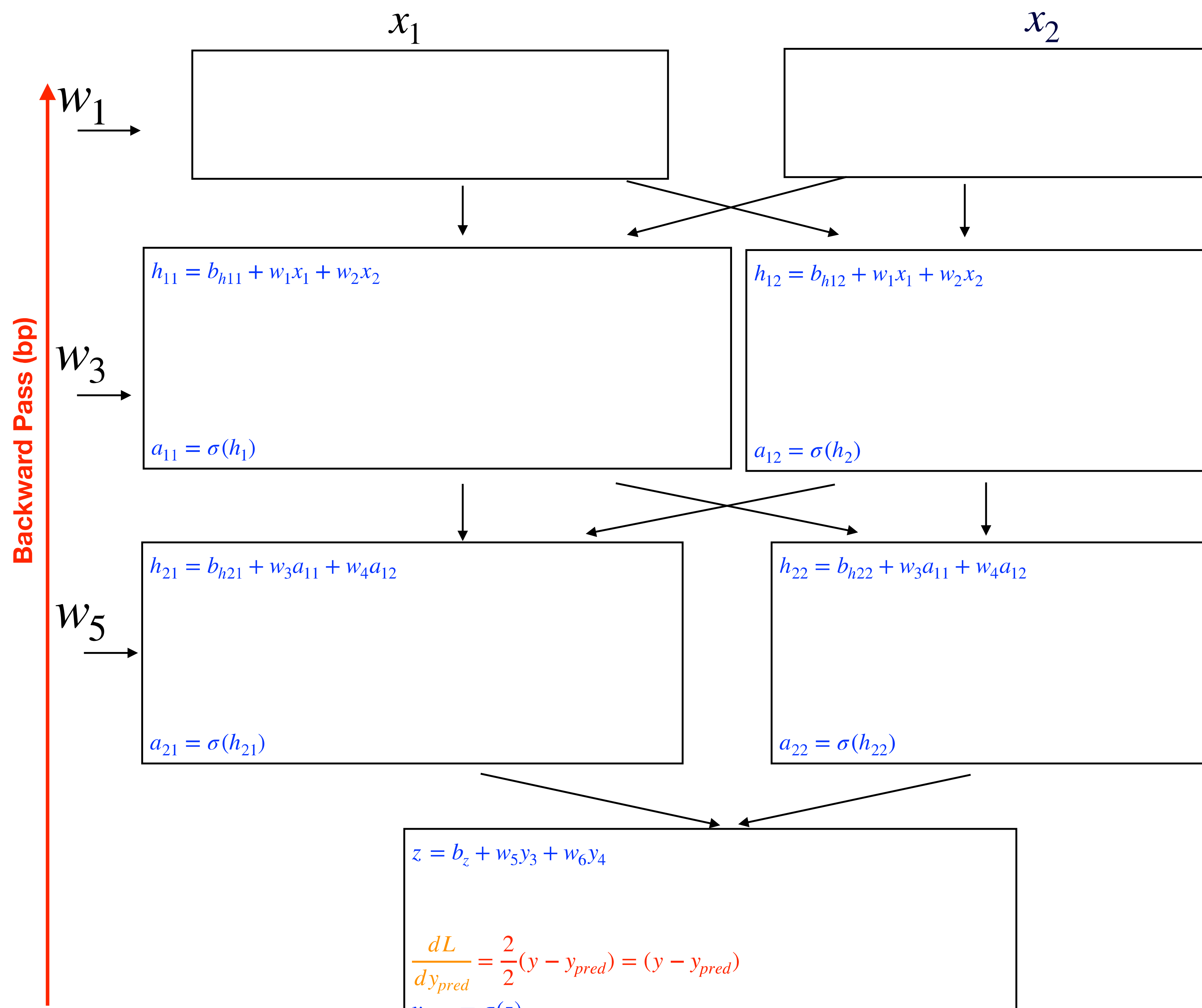


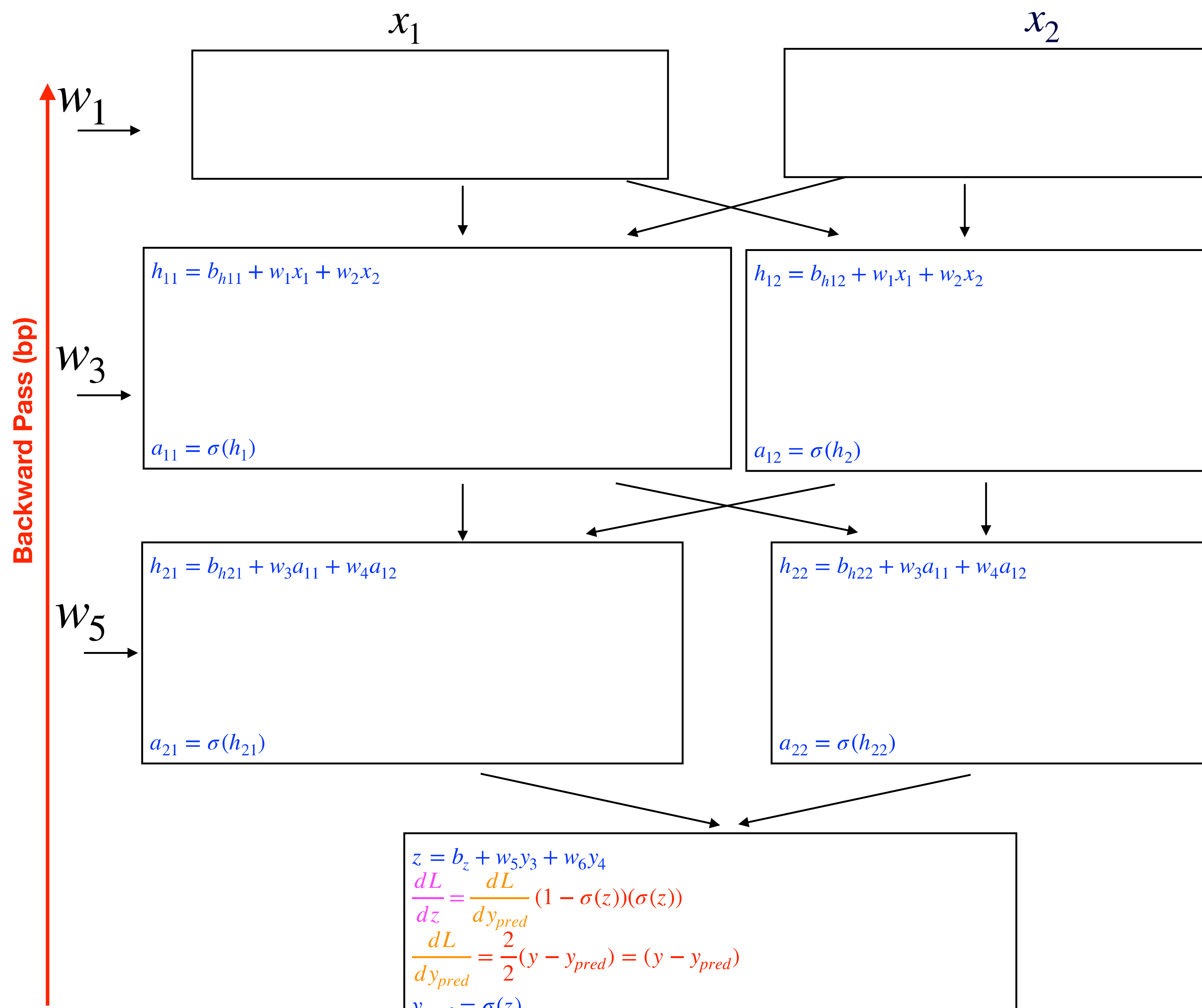
Backward Pass (bp)
w_1
w_3
w_5
x_1
x_2
$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial x_1} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial x_1} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}})\, w_1$
$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial w_1} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial w_1} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}})\, x_1$
$\frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial x_2} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial x_2} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}}$
$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial w_2} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial w_2} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{1}}$
$h_{11} = b_{h11} + w_1x_1 + w_2x_2$
$\frac{\partial L}{\partial h_{11}} = \frac{\partial L}{\partial a_{11}}\frac{\partial a_{11}}{\partial h_{11}} = \frac{\partial L}{\partial a_{11}}\sigma(h_{11})(1-\sigma(h_{11}))$
$\frac{\partial L}{\partial a_{11}} = \frac{\partial L}{\partial h_{21}}\frac{\partial h_{21}}{\partial a_{11}} + \frac{\partial L}{\partial h_{22}}\frac{\partial h_{22}}{\partial a_{11}} = (\frac{\partial L}{\partial h_{21}} + \frac{\partial L}{\partial h_{22}})\, w_3$
$a_{11} = \sigma(h_1)$
$h_{12} = b_{h12} + w_1x_1 + w_2x_2$
$\frac{\partial L}{\partial h_{12}} = \frac{\partial L}{\partial a_{12}}\frac{\partial a_{12}}{\partial h_{12}} = \frac{\partial L}{\partial a_{12}}\sigma(h_{12})(1-\sigma(h_{12}))$
$\frac{\partial L}{\partial a_{12}} = \frac{dL}{dh_{21}}\frac{\partial h_{21}}{\partial a_{12}} + \frac{\partial L}{\partial h_{22}}\frac{\partial h_{22}}{\partial a_{12}} = (\frac{dL}{dh_{21}} + \frac{\partial L}{\partial h_{22}})$
$a_{12} = \sigma(h_2)$
$h_{21} = b_{h21} + w_3a_{11} + w_4a_{12}$
$\frac{dL}{dh_{21}} = \frac{\partial L}{\partial a_{21}}\frac{da_{21}}{dh_{21}} = \frac{\partial L}{\partial a_{21}}(\sigma(h_{21})(1-\sigma(h_{21}))$
$\frac{\partial L}{\partial a_{21}} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial a_{21}} = \frac{\partial L}{\partial z} w_5$
$a_{21} = \sigma(h_{21})$
$h_{22} = b_{h22} + w_3a_{11} + w_4a_{12}$
$\frac{\partial L}{\partial h_{22}} = \frac{\partial L}{\partial a_{22}}\frac{da_{22}}{dh_{22}} = \frac{\partial L}{\partial a_{22}}(\sigma(h_{22})(1-\sigma(h_{22}))$
$\frac{\partial L}{\partial a_{22}} = \frac{dL}{dz}\frac{\partial z}{\partial a_{22}} = \frac{dL}{dz} w_6$
$a_{22} = \sigma(h_{22})$
$z = b_z + w_5y_3 + w_6y_4$
$\frac{dL}{dz} = \frac{dL}{dy_{pred}}(1-\sigma(z))(\sigma(z))$
$\frac{dL}{dy_{pred}} = \frac{2}{2}(y - y_{pred}) = (y - y_{pred})$

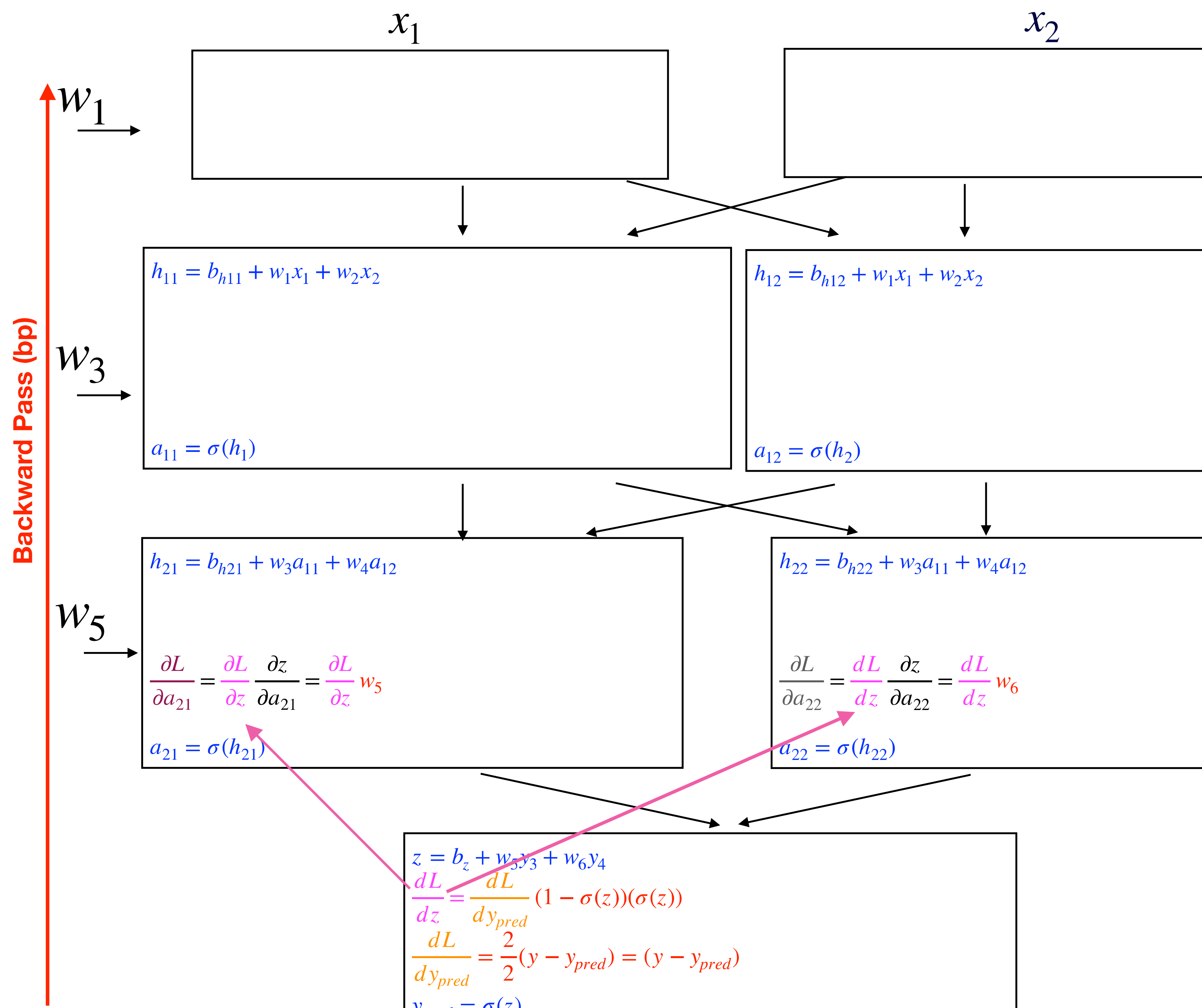


x_1
x_2
w_1
w_3
w_5
$h_{11} = b_{h11} + w_1x_1 + w_2x_2$
$a_{11} = \sigma(h_1)$
$h_{12} = b_{h12} + w_1x_1 + w_2x_2$
$a_{12} = \sigma(h_2)$
$h_{21} = b_{h21} + w_3a_{11} + w_4a_{12}$
$a_{21} = \sigma(h_{21})$
$h_{22} = b_{h22} + w_3a_{11} + w_4a_{12}$
$a_{22} = \sigma(h_{22})$
$z = b_z + w_5y_3 + w_6y_4$

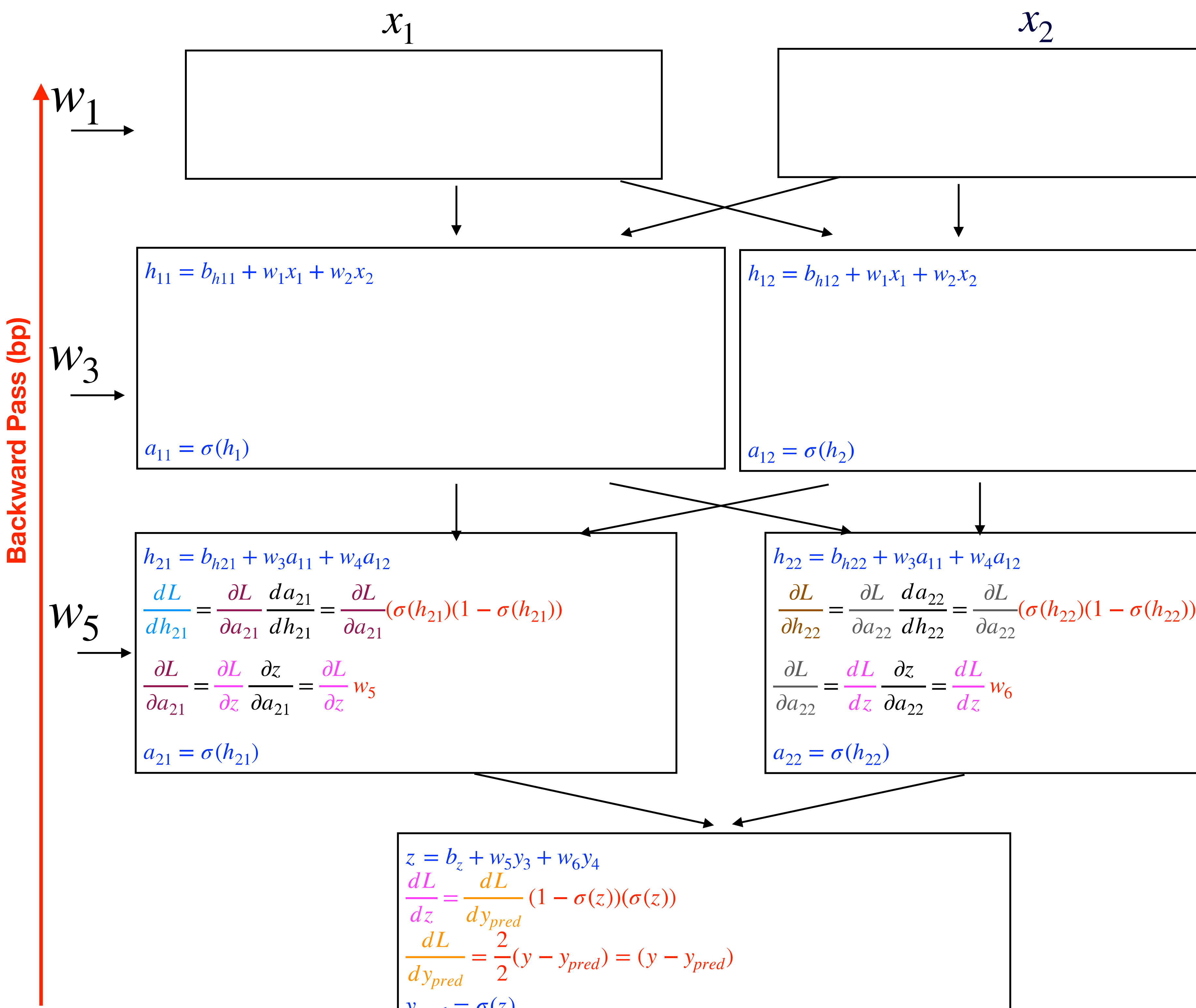


x_1
x_2
w_1
w_3
w_5
Backward Pass (bp)
$h_{11} = b_{h11} + w_1 x_1 + w_2 x_2$
$a_{11} = \sigma(h_1)$
$h_{12} = b_{h12} + w_1 x_1 + w_2 x_2$
$a_{12} = \sigma(h_2)$
$h_{21} = b_{h21} + w_3 a_{11} + w_4 a_{12}$
$a_{21} = \sigma(h_{21})$
$h_{22} = b_{h22} + w_3 a_{11} + w_4 a_{12}$
$a_{22} = \sigma(h_{22})$
$z = b_z + w_5 y_3 + w_6 y_4$
$\frac{dL}{dy_{pred}} = \frac{2}{2}(y - y_{pred}) = (y - y_{pred})$

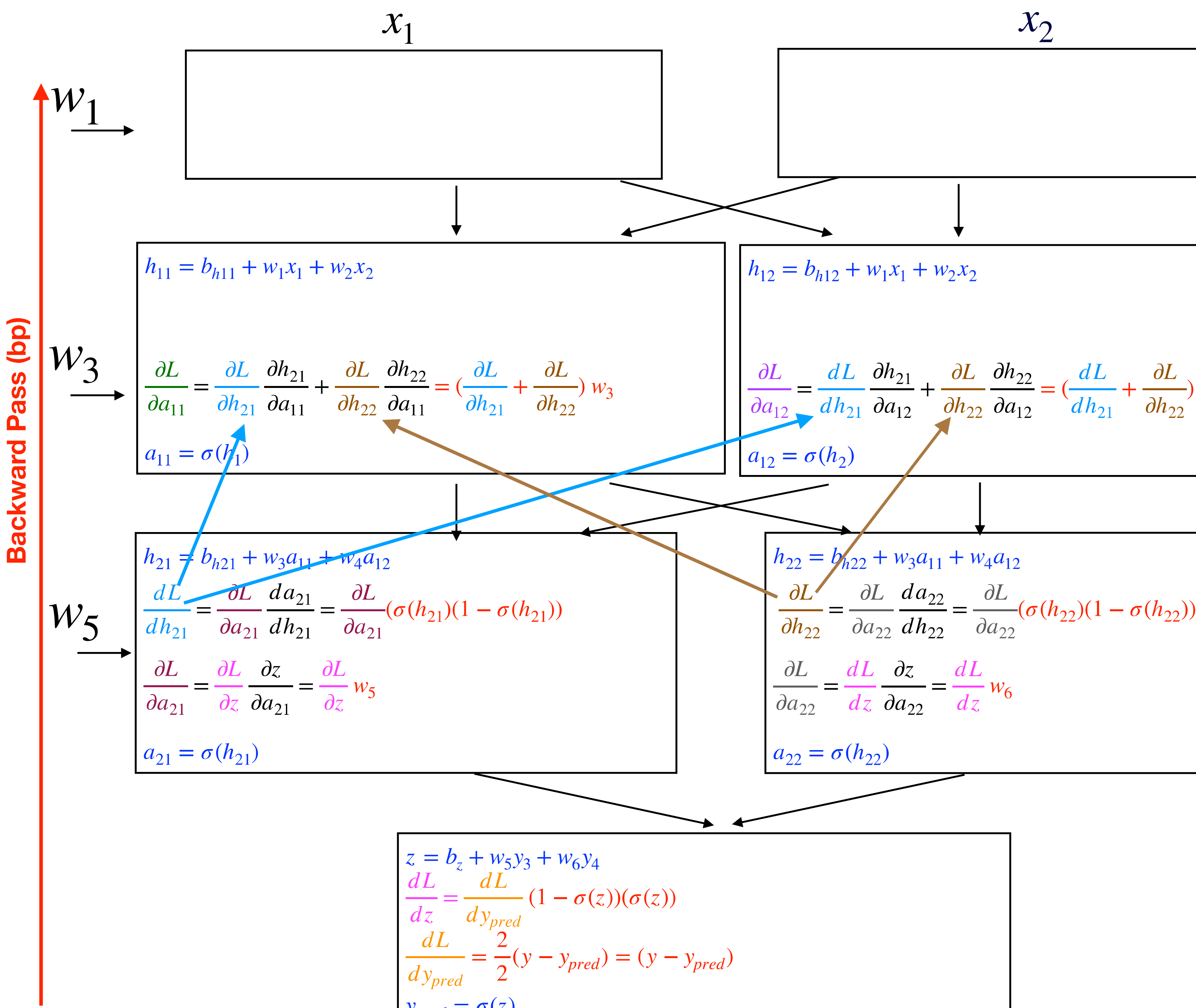

x_1
x_2
w_1
w_3
w_5
Backward Pass (bp)
$h_{11} = b_{h11} + w_1 x_1 + w_2 x_2$
$a_{11} = \sigma(h_1)$
$h_{12} = b_{h12} + w_1 x_1 + w_2 x_2$
$a_{12} = \sigma(h_2)$
$h_{21} = b_{h21} + w_3 a_{11} + w_4 a_{12}$
$a_{21} = \sigma(h_{21})$
$h_{22} = b_{h22} + w_3 a_{11} + w_4 a_{12}$
$a_{22} = \sigma(h_{22})$
$z = b_z + w_5 y_3 + w_6 y_4$
$\frac{dL}{dz} = \frac{dL}{dy_{pred}}(1 - \sigma(z))(\sigma(z))$
$\frac{dL}{dy_{pred}} = \frac{2}{2}(y - y_{pred}) = (y - y_{pred})$

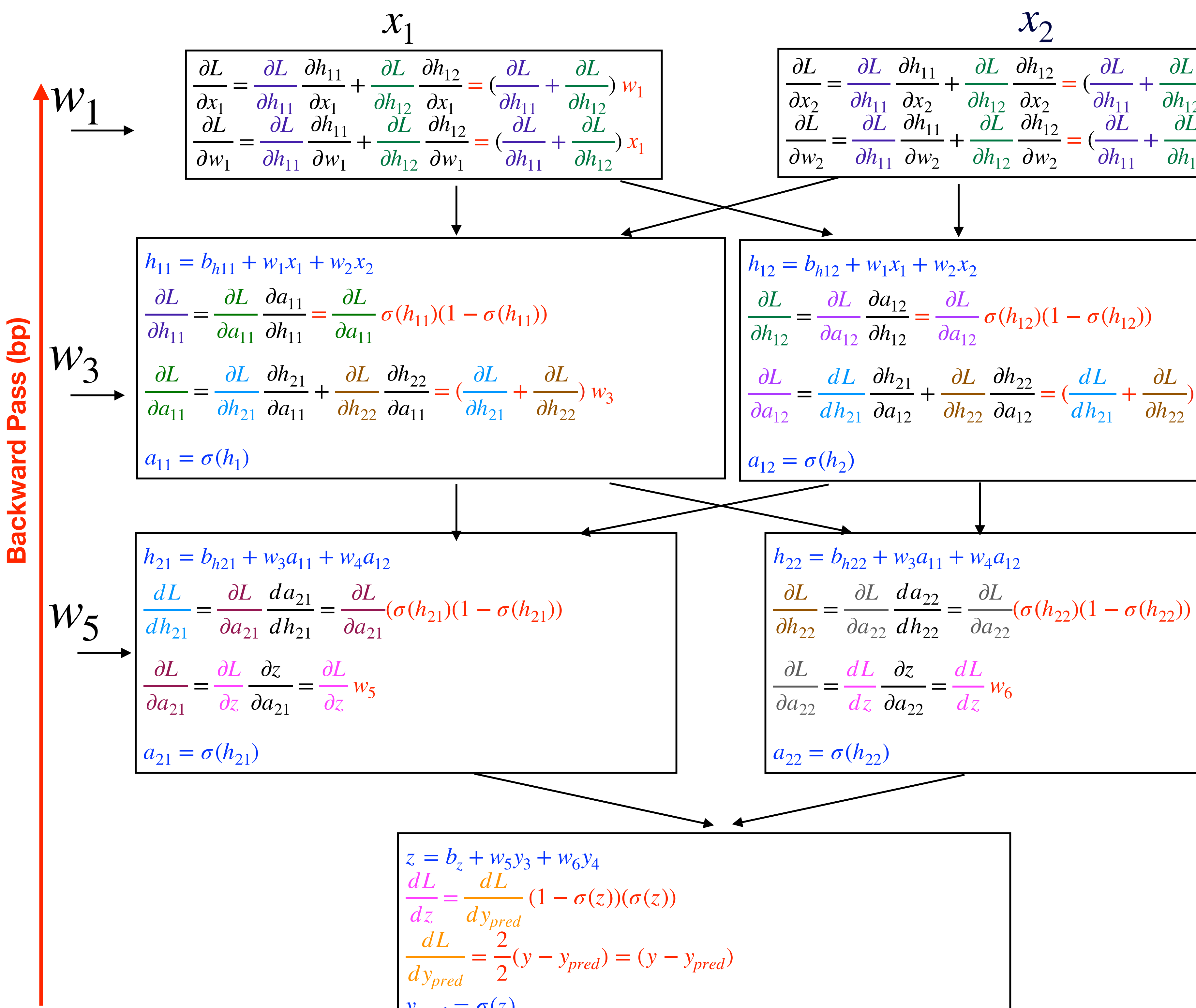


Backward Pass (bp)
w_1
w_3
w_5
x_1
x_2
$h_{11} = b_{h11} + w_1 x_1 + w_2 x_2$
$a_{11} = \sigma(h_1)$
$h_{12} = b_{h12} + w_1 x_1 + w_2 x_2$
$a_{12} = \sigma(h_2)$
$h_{21} = b_{h21} + w_3 a_{11} + w_4 a_{12}$
$\frac{\partial L}{\partial a_{21}} = \frac{\partial L}{\partial z} \frac{\partial z}{\partial a_{21}} = \frac{\partial L}{\partial z} w_5$
$a_{21} = \sigma(h_{21})$
$h_{22} = b_{h22} + w_3 a_{11} + w_4 a_{12}$
$\frac{\partial L}{\partial a_{22}} = \frac{dL}{dz} \frac{\partial z}{\partial a_{22}} = \frac{dL}{dz} w_6$
$a_{22} = \sigma(h_{22})$
$z = b_z + w_5 y_3 + w_6 y_4$
$\frac{dL}{dz} = \frac{dL}{dy_{pred}} (1 - \sigma(z))(\sigma(z))$
$\frac{dL}{dy_{pred}} = \frac{2}{2}(y - y_{pred}) = (y - y_{pred})$

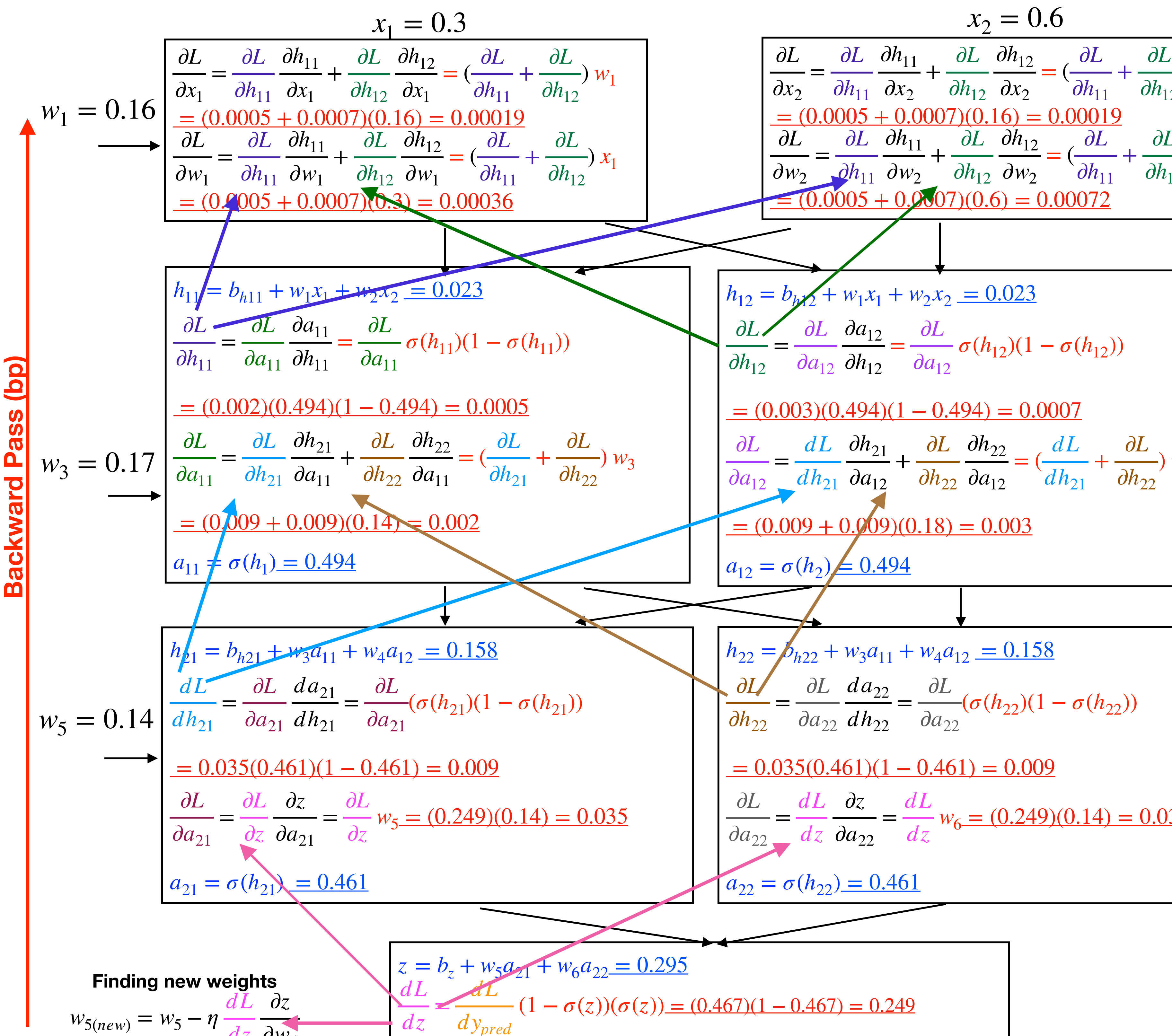


x_1
x_2
Backward Pass (bp)
w_1
w_3
w_5
$h_{11} = b_{h11} + w_1x_1 + w_2x_2$
$a_{11} = \sigma(h_1)$
$h_{12} = b_{h12} + w_1x_1 + w_2x_2$
$a_{12} = \sigma(h_2)$
$h_{21} = b_{h21} + w_3a_{11} + w_4a_{12}$
$\frac{dL}{dh_{21}} = \frac{\partial L}{\partial a_{21}}\frac{da_{21}}{dh_{21}} = \frac{\partial L}{\partial a_{21}}(\sigma(h_{21})(1-\sigma(h_{21}))$
$\frac{\partial L}{\partial a_{21}} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial a_{21}} = \frac{\partial L}{\partial z}w_5$
$a_{21} = \sigma(h_{21})$
$h_{22} = b_{h22} + w_3a_{11} + w_4a_{12}$
$\frac{\partial L}{\partial h_{22}} = \frac{\partial L}{\partial a_{22}}\frac{da_{22}}{dh_{22}} = \frac{\partial L}{\partial a_{22}}(\sigma(h_{22})(1-\sigma(h_{22}))$
$\frac{\partial L}{\partial a_{22}} = \frac{dL}{dz}\frac{\partial z}{\partial a_{22}} = \frac{dL}{dz}w_6$
$a_{22} = \sigma(h_{22})$
$z = b_z + w_5y_3 + w_6y_4$
$\frac{dL}{dz} = \frac{dL}{dy_{pred}}(1-\sigma(z))(\sigma(z))$
$\frac{dL}{dy_{pred}} = \frac{2}{2}(y - y_{pred}) = (y - y_{pred})$


Backward Pass (bp)
w_1
w_3
w_5
x_1
x_2
$h_{11} = b_{h11} + w_1x_1 + w_2x_2$
$\frac{\partial L}{\partial a_{11}} = \frac{\partial L}{\partial h_{21}} \frac{\partial h_{21}}{\partial a_{11}} + \frac{\partial L}{\partial h_{22}} \frac{\partial h_{22}}{\partial a_{11}} = (\frac{\partial L}{\partial h_{21}} + \frac{\partial L}{\partial h_{22}})\ w_3$
$a_{11} = \sigma(h_1)$
$h_{12} = b_{h12} + w_1x_1 + w_2x_2$
$\frac{\partial L}{\partial a_{12}} = \frac{dL}{dh_{21}} \frac{\partial h_{21}}{\partial a_{12}} + \frac{\partial L}{\partial h_{22}} \frac{\partial h_{22}}{\partial a_{12}} = (\frac{dL}{dh_{21}} + \frac{\partial L}{\partial h_{22}})$
$a_{12} = \sigma(h_2)$
$h_{21} = b_{h21} + w_3a_{11} + w_4a_{12}$
$\frac{dL}{dh_{21}} = \frac{\partial L}{\partial a_{21}} \frac{da_{21}}{dh_{21}} = \frac{\partial L}{\partial a_{21}}(\sigma(h_{21})(1-\sigma(h_{21}))$
$\frac{\partial L}{\partial a_{21}} = \frac{\partial L}{\partial z} \frac{\partial z}{\partial a_{21}} = \frac{\partial L}{\partial z} w_5$
$a_{21} = \sigma(h_{21})$
$h_{22} = b_{h22} + w_3a_{11} + w_4a_{12}$
$\frac{\partial L}{\partial h_{22}} = \frac{\partial L}{\partial a_{22}} \frac{da_{22}}{dh_{22}} = \frac{\partial L}{\partial a_{22}}(\sigma(h_{22})(1-\sigma(h_{22}))$
$\frac{\partial L}{\partial a_{22}} = \frac{dL}{dz} \frac{\partial z}{\partial a_{22}} = \frac{dL}{dz} w_6$
$a_{22} = \sigma(h_{22})$
$z = b_z + w_5y_3 + w_6y_4$
$\frac{dL}{dz} = \frac{dL}{dy_{pred}} (1-\sigma(z))(\sigma(z))$
$\frac{dL}{dy_{pred}} = \frac{2}{2}(y - y_{pred}) = (y - y_{pred})$



Backward Pass (bp)
w_1
w_3
w_5
x_1
x_2
$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial x_1} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial x_1} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}})\, w_1$
$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial w_1} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial w_1} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}})\, x_1$
$\frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial x_2} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial x_2} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}}$
$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial w_2} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial w_2} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{1}}$
$h_{11} = b_{h11} + w_1x_1 + w_2x_2$
$\frac{\partial L}{\partial h_{11}} = \frac{\partial L}{\partial a_{11}}\frac{\partial a_{11}}{\partial h_{11}} = \frac{\partial L}{\partial a_{11}}\sigma(h_{11})(1-\sigma(h_{11}))$
$\frac{\partial L}{\partial a_{11}} = \frac{\partial L}{\partial h_{21}}\frac{\partial h_{21}}{\partial a_{11}} + \frac{\partial L}{\partial h_{22}}\frac{\partial h_{22}}{\partial a_{11}} = (\frac{\partial L}{\partial h_{21}} + \frac{\partial L}{\partial h_{22}})\, w_3$
$a_{11} = \sigma(h_1)$
$h_{12} = b_{h12} + w_1x_1 + w_2x_2$
$\frac{\partial L}{\partial h_{12}} = \frac{\partial L}{\partial a_{12}}\frac{\partial a_{12}}{\partial h_{12}} = \frac{\partial L}{\partial a_{12}}\sigma(h_{12})(1-\sigma(h_{12}))$
$\frac{\partial L}{\partial a_{12}} = \frac{dL}{dh_{21}}\frac{\partial h_{21}}{\partial a_{12}} + \frac{\partial L}{\partial h_{22}}\frac{\partial h_{22}}{\partial a_{12}} = (\frac{dL}{dh_{21}} + \frac{\partial L}{\partial h_{22}})$
$a_{12} = \sigma(h_2)$
$h_{21} = b_{h21} + w_3a_{11} + w_4a_{12}$
$\frac{dL}{dh_{21}} = \frac{\partial L}{\partial a_{21}}\frac{da_{21}}{dh_{21}} = \frac{\partial L}{\partial a_{21}}(\sigma(h_{21})(1-\sigma(h_{21}))$
$\frac{\partial L}{\partial a_{21}} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial a_{21}} = \frac{\partial L}{\partial z}w_5$
$a_{21} = \sigma(h_{21})$
$h_{22} = b_{h22} + w_3a_{11} + w_4a_{12}$
$\frac{\partial L}{\partial h_{22}} = \frac{\partial L}{\partial a_{22}}\frac{da_{22}}{dh_{22}} = \frac{\partial L}{\partial a_{22}}(\sigma(h_{22})(1-\sigma(h_{22}))$
$\frac{\partial L}{\partial a_{22}} = \frac{dL}{dz}\frac{\partial z}{\partial a_{22}} = \frac{dL}{dz}w_6$
$a_{22} = \sigma(h_{22})$
$z = b_z + w_5y_3 + w_6y_4$
$\frac{dL}{dz} = \frac{dL}{dy_{pred}}(1-\sigma(z))(\sigma(z))$
$\frac{dL}{dy_{pred}} = \frac{2}{2}(y - y_{pred}) = (y - y_{pred})$



Backward Pass (bp)
$x_1 = 0.3$
$x_2 = 0.6$
$w_1 = 0.16$
$w_3 = 0.17$
$w_5 = 0.14$
$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial x_1} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial x_1} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}})\, w_1$
$= (0.0005 + 0.0007)(0.16) = 0.00019$
$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial w_1} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial w_1} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}})\, x_1$
$= (0.0005 + 0.0007)(0.3) = 0.00036$
$\frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial x_2} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial x_2} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{12}}$
$= (0.0005 + 0.0007)(0.16) = 0.00019$
$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial h_{11}}\frac{\partial h_{11}}{\partial w_2} + \frac{\partial L}{\partial h_{12}}\frac{\partial h_{12}}{\partial w_2} = (\frac{\partial L}{\partial h_{11}} + \frac{\partial L}{\partial h_{1}}$
$= (0.0005 + 0.0007)(0.6) = 0.00072$
$h_{11} = b_{h11} + w_1x_1 + w_2x_2 = 0.023$
$\frac{\partial L}{\partial h_{11}} = \frac{\partial L}{\partial a_{11}}\frac{\partial a_{11}}{\partial h_{11}} = \frac{\partial L}{\partial a_{11}}\sigma(h_{11})(1 - \sigma(h_{11}))$
$= (0.002)(0.494)(1 - 0.494) = 0.0005$
$\frac{\partial L}{\partial a_{11}} = \frac{\partial L}{\partial h_{21}}\frac{\partial h_{21}}{\partial a_{11}} + \frac{\partial L}{\partial h_{22}}\frac{\partial h_{22}}{\partial a_{11}} = (\frac{\partial L}{\partial h_{21}} + \frac{\partial L}{\partial h_{22}})\, w_3$
$= (0.009 + 0.009)(0.14) = 0.002$
$a_{11} = \sigma(h_1) = 0.494$
$h_{12} = b_{h12} + w_1x_1 + w_2x_2 = 0.023$
$\frac{\partial L}{\partial h_{12}} = \frac{\partial L}{\partial a_{12}}\frac{\partial a_{12}}{\partial h_{12}} = \frac{\partial L}{\partial a_{12}}\sigma(h_{12})(1 - \sigma(h_{12}))$
$= (0.003)(0.494)(1 - 0.494) = 0.0007$
$\frac{\partial L}{\partial a_{12}} = \frac{dL}{dh_{21}}\frac{\partial h_{21}}{\partial a_{12}} + \frac{\partial L}{\partial h_{22}}\frac{\partial h_{22}}{\partial a_{12}} = (\frac{dL}{dh_{21}} + \frac{\partial L}{\partial h_{22}})$
$= (0.009 + 0.009)(0.18) = 0.003$
$a_{12} = \sigma(h_2) = 0.494$
$h_{21} = b_{h21} + w_3a_{11} + w_4a_{12} = 0.158$
$\frac{dL}{dh_{21}} = \frac{\partial L}{\partial a_{21}}\frac{da_{21}}{dh_{21}} = \frac{\partial L}{\partial a_{21}}(\sigma(h_{21})(1 - \sigma(h_{21}))$
$= 0.035(0.461)(1 - 0.461) = 0.009$
$\frac{\partial L}{\partial a_{21}} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial a_{21}} = \frac{\partial L}{\partial z} w_5 = (0.249)(0.14) = 0.035$
$a_{21} = \sigma(h_{21}) = 0.461$
$h_{22} = b_{h22} + w_3a_{11} + w_4a_{12} = 0.158$
$\frac{\partial L}{\partial h_{22}} = \frac{\partial L}{\partial a_{22}}\frac{da_{22}}{dh_{22}} = \frac{\partial L}{\partial a_{22}}(\sigma(h_{22})(1 - \sigma(h_{22}))$
$= 0.035(0.461)(1 - 0.461) = 0.009$
$\frac{\partial L}{\partial a_{22}} = \frac{dL}{dz}\frac{\partial z}{\partial a_{22}} = \frac{dL}{dz} w_6 = (0.249)(0.14) = 0.0$
$a_{22} = \sigma(h_{22}) = 0.461$
$z = b_z + w_5a_{21} + w_6a_{22} = 0.295$
$\frac{dL}{dz} = \frac{dL}{dy_{pred}}(1 - \sigma(z))(\sigma(z)) = (0.467)(1 - 0.467) = 0.249$
Finding new weights
$w_{5(new)} = w_5 - \eta \frac{dL}{dz}\frac{\partial z}{\partial w_5}$

Backpropagation- forward pass and Gradient calculations

- As seen in the previous slide we perform the forward pass to
    - linear combination(or a function) of inputs
    - sigmoid (activation function) of the above function
- We use the value from the output layer to predict the value of
- we calculate loss function Using MSE or Cross entropy

Backpropagation- backward pass

- during backward pass we propagate the gradient through all nodes
  - first, we take a sum of products of all incoming gradient from the next nodes and weights of current noc
    - $\sum_{i=1}^{NodesInNextLayer} \frac{\partial L}{\partial h_{layer,i}} w_{layer})$
    - e.g. $(\frac{\partial L}{\partial h_{21}} + \frac{\partial L}{\partial h_{22}})\, w_3$
      - where h_{21}, h_{22} are nodes in the next layer in MLP connect to current node h_{11}
      - w_3 is the weight of the current layer h_{11}
  - multiply this sum with the derivative of the sigmoid function

Backpropagation - weight calculations

- using the new gradient we find the new weights as
  $w_{new} = w_{old} - \eta(\frac{dL}{dNode}\sigma(Node))$
  - $\frac{dL}{dNode}$ the node gets from the previous node during back
  - $\sigma(Node)$ is its own output
  - w_{old} is the node's current weight