Weight Updates with ReLU Hidden, sigmoid output and binary cross-entropy

- Training Input: (x1,x2)=(1,1)
- Target Output: d=0.0

z_1^1 = 0.25*1+0.75*1=1.0

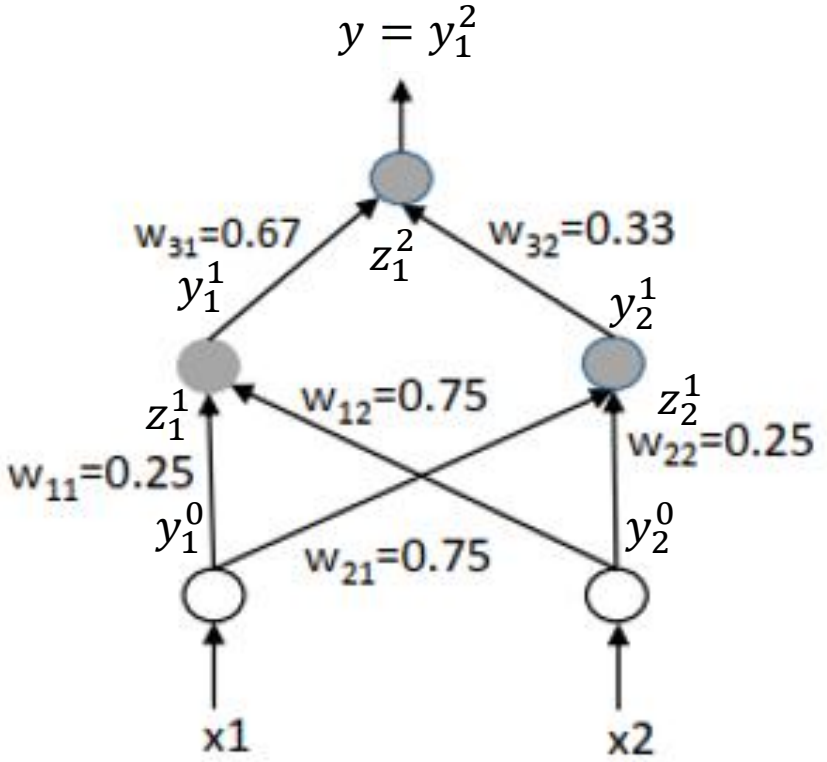
z_2^1 = 0.75*1+0.25*1=1.0

y_1^1 = ReLU(z_1^1)= 1 y_2^1 = ReLU(z_2^1)= 1

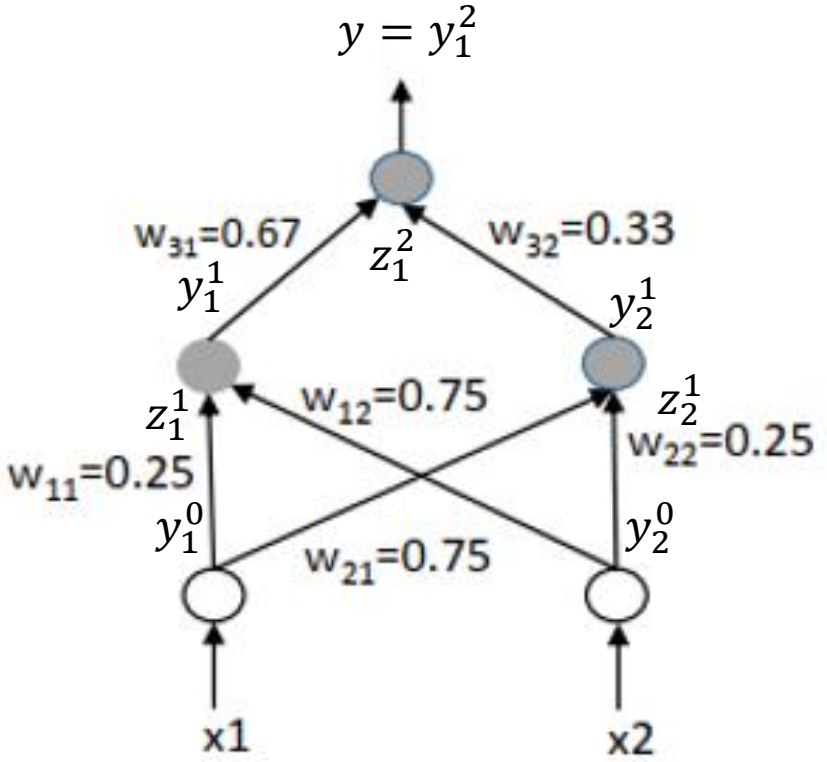
z_1^2 = 0.67*1+0.33*1=1.0

y_1^2 = sigmoid(z_1^2)=e/(1+e)

div = -d*log(y)-(1-d)*log(1-y)



Weight Updates with ReLU Hidden, sigmoid output and binary cross-entropy



z_1^1 = 0.25*1+0.75*1=1.0

z_2^1 = 0.75*1+0.25*1=1.0

y_1^1 = ReLU(z_1^1)= 1 y_2^1 = ReLU(z_2^1)= 1

z_1^2 = 0.67*1+0.33*1=1.0

y_1^2 = sigmoid(z_1^2)=e/(1+e)

div = -d*log(y)-(1-d)*log(1-y)

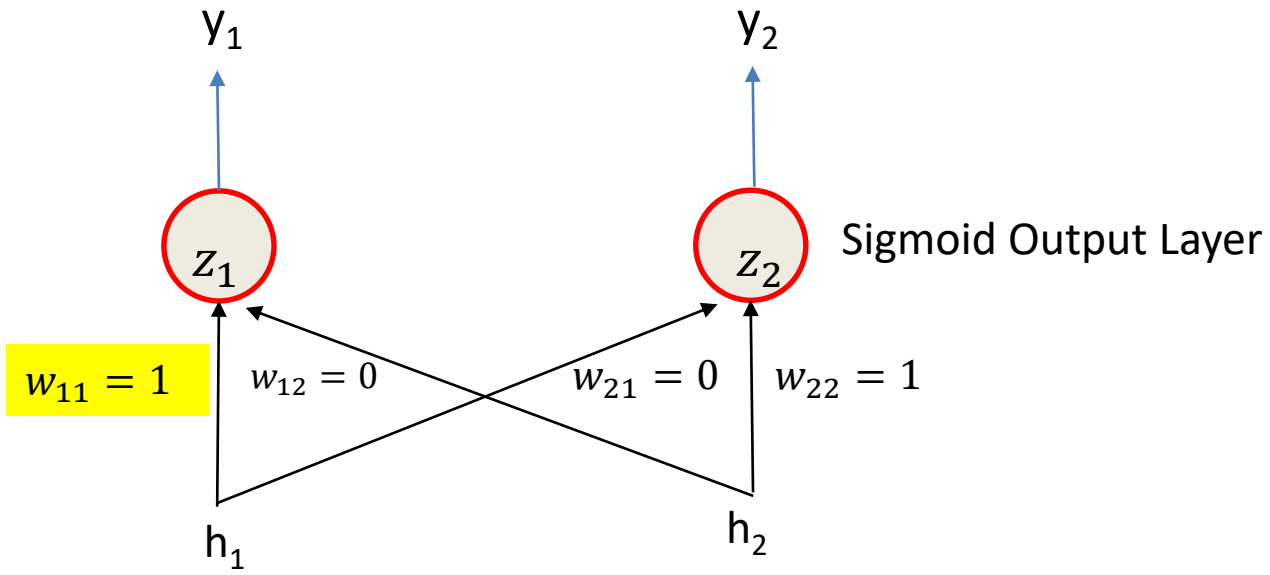
$\frac{\delta div}{\delta y}$=-d/y+(1-d)/(1-y)=1/(1-y)= (1+e)

$\frac{\delta div}{\delta z_1^2} = \frac{\delta Div}{\delta y_1^2} * \frac{\delta y_1^2}{\delta z_1^2}$ = (1+e)*sigmoid(z_1^2)*(1-sigmoid(z_1^2))=e/(1+e)

$\frac{\delta div}{\delta w_{31}} = \frac{\delta Div}{\delta z_1^2} * \frac{\delta z_1^2}{\delta w_{31}} = e/(1+e) * y_1^1$=e/(1+e)

$$w_{31} = w_{31} - \eta * \frac{\delta div}{\delta w_{31}}$$=0.67-0.1*e/(1+e)=0.597

Weight Updates with Sigmoid Output



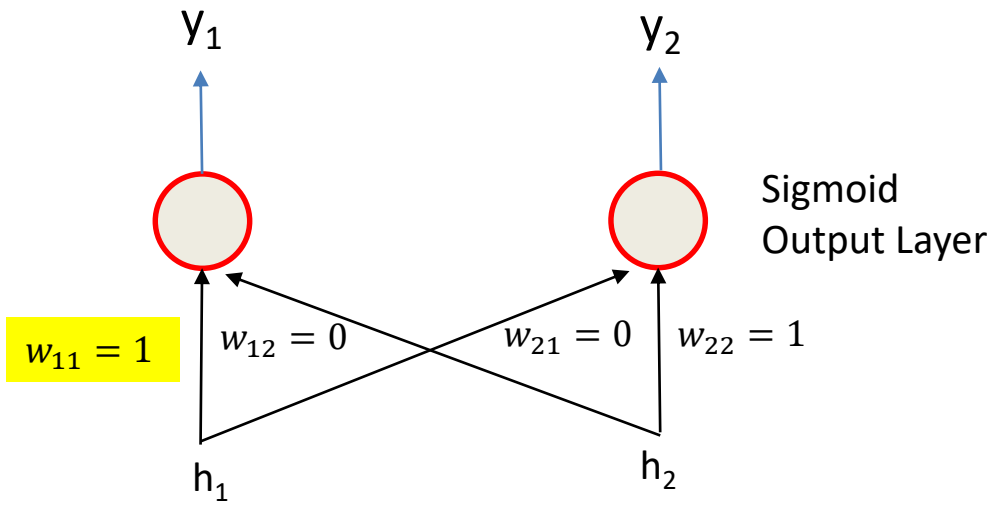
- Training input: $(h_1, h_2) = (1, 0)$ target output: $(d_1,d_2)=(0,1)$
- Activation: sigmoid, Bias=0
- *div* function: cross-entropy
- Learning rate: 0.1
- What will be the value of w_{11} in next iteration?

- Input to output node $z_1 = w_{11}=1$; $z_2 =0$
- $y_1 = 1 / (1+e^{-1}) = e/(1+e)$ $y_2 = 1/ (1+e^{0})=1/2$
- *div* = $-d_1 \log(y_1) - d_2 \log(y_2) = -\log(y_2) = \log 2$

Weight Updates with Sigmoid Outputs …

w_{11}(t+1)=?

Input (h1, h2)=(1.0, 0.0)
target output: (d_1,d_2)=(0,1)

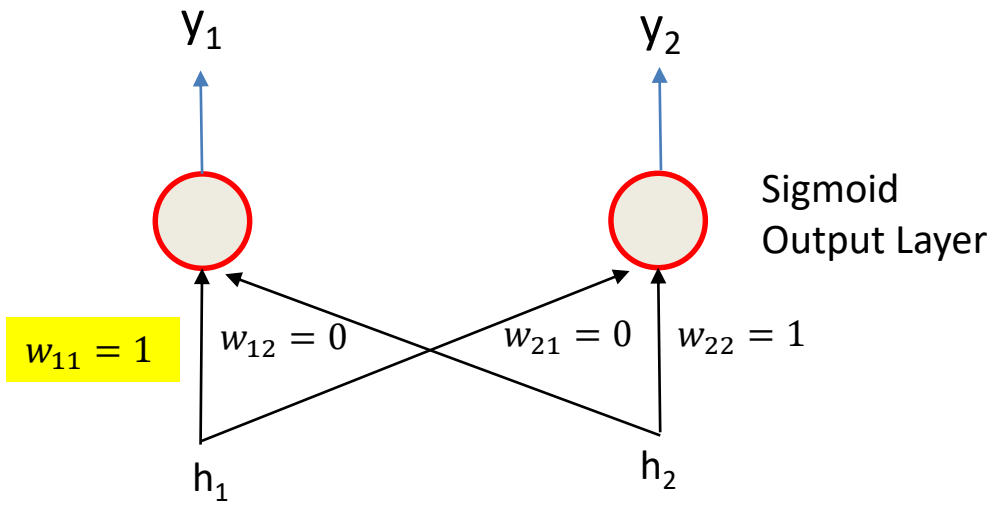


Weight Updates with Sigmoid Outputs …

Input (h1, h2)=(1.0, 0.0)
target output: $(d_1,d_2)=(0,1)$

$w_{11}(t+1)=?$

$div = -d_1 \log(y_1) - d_2 \log(y_2)$



$$\frac{\delta div}{\delta y_1} = -\frac{d_1}{y_1}$$

$$\frac{\delta y_1}{\delta z_1} = \frac{1}{1+e^{-1}} * \left(1 - \frac{1}{1+e^{-1}}\right) = \frac{e}{(1+e)^2}$$

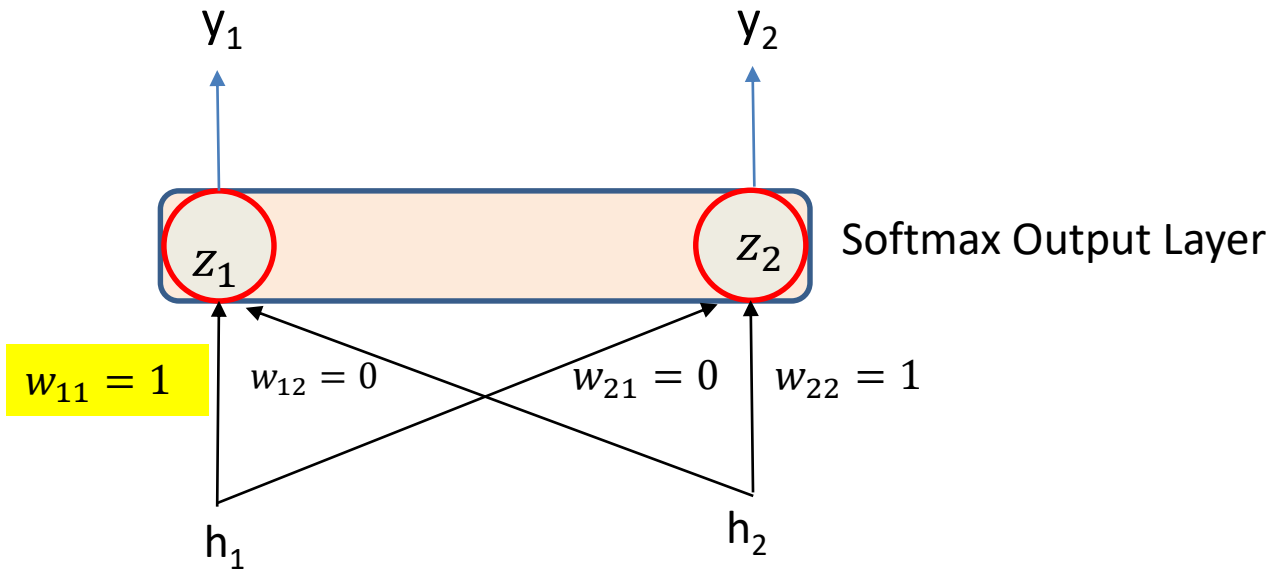
$$w_{11}(t+1) = w_{11}(t) - \eta \frac{\delta div}{\delta w_{11}}$$

$$= w_{11}(t) - 0.1 * \frac{\delta div}{\delta y_1}\frac{\delta y_1}{\delta z_1}\frac{\delta z_1}{\delta w_{11}}$$

$$= w_{11}(t) - 0.1 * 0 * \frac{e}{(1+e)^2} * h_1$$

$$= 1.0$$

Weight Updates with Softmax



- Training input: $(h_1, h_2) = (1, 0)$ target output: $(d_1,d_2)=(0,1)$
- *div* function: cross-entropy
- Learning rate: 0.1
- Bias = 0
- What will be the value of w_{11} in next iteration?

- Input to softmax node $z_1 = w_{11}=1$; $z_2 =0$
- $y_1 = e / (1+e)$
- $y_2 = 1/ (1+e)$
- *div* = $-d_1 \log(y_1) - d_2 \log(y_2) = -\log(y_2) = \log(1+e) = 1.313$

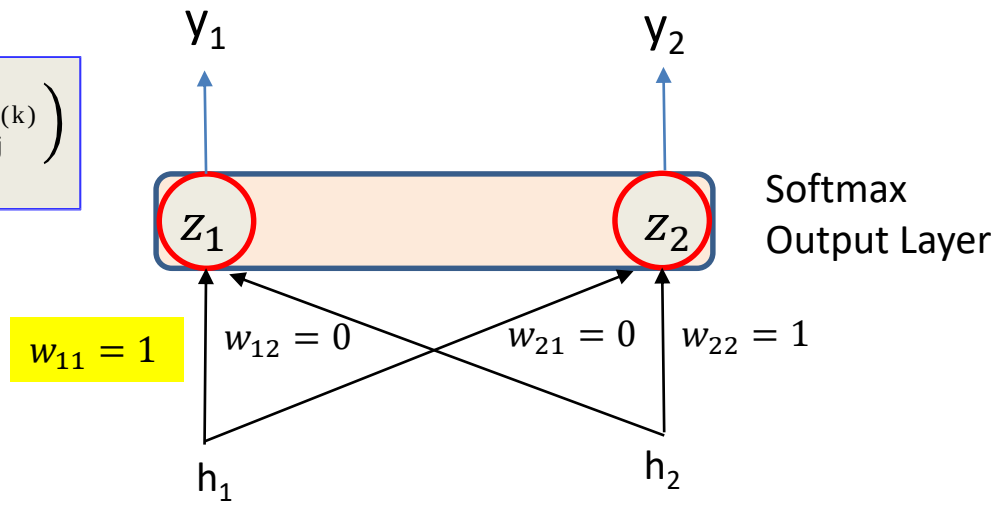
Weight Updates with Softmax …

$$\frac{\partial Div}{\partial z_i^{(k)}} = \sum_j \frac{\partial Div}{\partial y_j^{(k)}} y_i^{(k)} \left(\delta_{ij} - y_j^{(k)}\right)$$

$$w_{11}(t+1) = ?$$

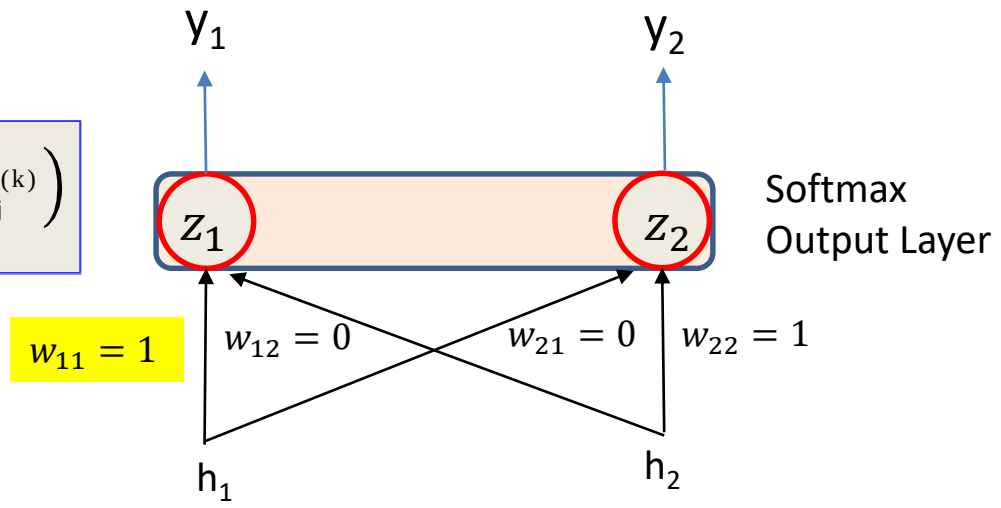
Training input: (h_1, h_2) = (1, 0)

Target output: (d_1,d_2)=(0,1)



Weight Updates with Softmax …

$$\frac{\partial Div}{\partial z_i^{(k)}} = -\sum_j \frac{\partial Div}{\partial y_j^{(k)}} y_i^{(k)} \left(\delta_{ij} - y_j^{(k)}\right)$$



- Change in w_{11} = δw_{11} = -0.1 * d*div*/dw_{11}
  = -0.1 * d*div*/dz_1 * dz_1/dw_{11} = -0.1 * d*div*/dz_1 * h_1

- d*div*/dy_1 = -d_1/y_1 = 0 d*div*/dy_2=-d_2/y_2 = -1/y_2
- d*div*/dz_1 = d*div*/dy_1*y_1(1-y_1) + d*div*/dy_2*y_1(-y_2) = - y_1 y_2 / y_2 = -y_1

- Input to softmax node z_1 = w_{11}=1; z_2 =0 ; y_1 = e / (1+e)
- δw_{11} = −0.1 * e/(1+e) = -0.0731