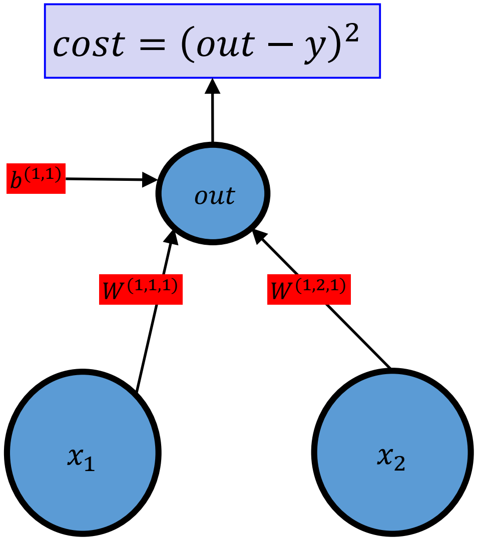
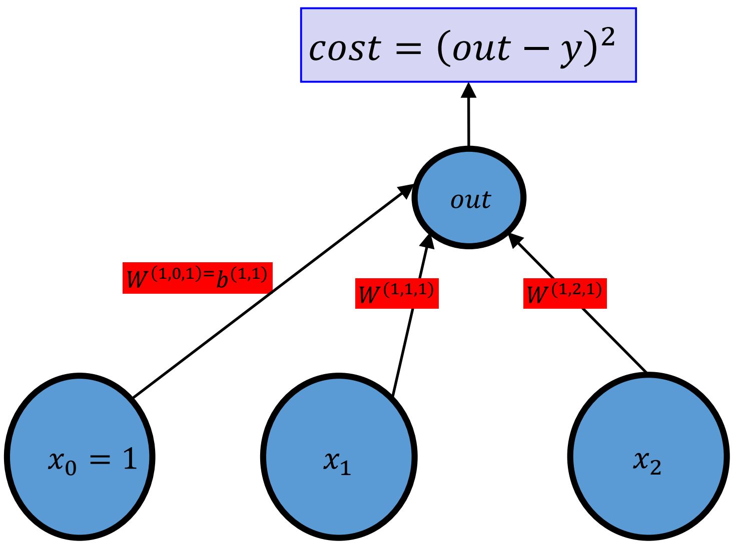


$$out = b^{(1,1)} + W^{(1,1,1)}x_1 + W^{(1,2,1)}x_2$$

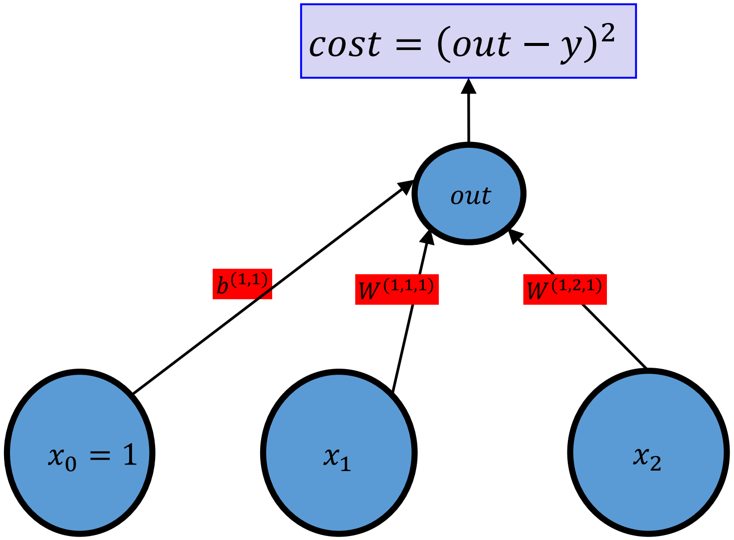


$$out = b^{(1,1)} + W^{(1,1,1)} x_1 + W^{(1,2,1)} x_2$$
$$= W^{(1,0,1)} x_0 + W^{(1,1,1)} x_1 + W^{(1,2,1)} x_2$$

$$= \sum_i W^{(1,i,1)} x_i$$

Note: for the identity activation function, $\frac{\partial out}{\partial sum} = 1$ since $out = sum$

Total Cost



- $cost_{total} = \sum_i \ cost(out^{(i)} - y^{(i)})$
  $$= \sum_i (out^{(i)} - y^{(i)})^2$$
  $$= \sum_i (\sum_j W^{(1,j,1)} x_j^{(i)} - y^{(i)})^2$$
- $\frac{\partial cost_{total}}{\partial W^{(1,j,1)}} = \sum_i \ 2x_j^{(i)} (\sum_j \ W^{(1,j,1)} x_j^{(i)} - y^{(i)}) =$
  $$2 \sum_i x_j^{(i)} (W^T x^{(i)} - y^{(i)})$$

- $\frac{\partial cost_{total}}{\partial W^{(1,j,1)}} = \sum_i \ 2x_j^{(i)}(W^T x^{(i)} - y^{(i)})$

- $\frac{\partial cost_{total}}{\partial W^{(1,*,1)}} =$
$$2sum\left(\begin{bmatrix} x_1^{(1)}(W^T x^{(1)} - y^{(1)}) & \dots & x_j^{(i)}(W^T x^{(i)} - y^{(i)}) & & \\ \dots & \dots & \dots & \cdots & \\ x_j^{(1)}(W^T x^{(1)} - y^{(1)}) & \dots & x_j^{(i)}\ (W^T x_j^{(i)} - y^{(i)}) & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & \end{bmatrix}, 1\right)$$

- $$2sum\left(\begin{bmatrix} x_1^{(1)}(W^T x^{(1)} - y^{(1)}) & \dots & x_1^{(i)}(W^T x^{(i)} - y^{(i)}) & & \\ \dots & \dots & \dots & \cdots & \\ x_j^{(1)}(W^T x^{(1)} - y^{(1)}) & \dots & x_j^{(i)}\,(W^T x^{(i)} - y^{(i)}) & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & \end{bmatrix}, 1\right)$$

$$= 2sum\left(\left(\begin{bmatrix} (W^T x^{(1)} - y^{(1)}) & \dots & (W^T x^{(i)} - y^{(i)}) & & \\ \dots & \dots & \dots & \cdots & \\ (W^T x^{(1)} - y^{(1)}) & \dots & W^T x^{(i)} - y^{(i)}) & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & \end{bmatrix}\right) * \begin{bmatrix} x_1^{(1)} & \dots & x_j^{(i)} & & \\ \dots & \dots & \dots & \cdots & \\ x_1^{(1)} & \dots & x_j^{(i)} & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & \end{bmatrix}, 1\right)$$

$$= sum((W^T x - y) * \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & & \\ x_2^{(1)} & x_2^{(2)} & \dots & \cdots & \\ \dots & \dots & \dots & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & \end{bmatrix}, 1)$$

$$\frac{\partial cost_{total}}{\partial W^{(1,j,1)}} = \sum_{i} 2x_j^{(i)}(W^T x^{(i)} - y^{(i)})$$

$$x = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & & \\ x_2^{(1)} & x_2^{(2)} & \cdots & \cdots & \\ \cdots & \cdots & \cdots & & \\ & \vdots & & \ddots & \vdots \\ & & & \cdots & \end{bmatrix}, (W^T x - y) = \begin{bmatrix} W^T x^{(1)} - y^{(1)} \\ \cdots \\ W^T x^{(i)} - y^{(i)} \\ \cdots \end{bmatrix}$$

$$\frac{\partial cost_{total}}{\partial W^{(1,j,1)}} = x(W^T x - y)$$