net
loss
view   output

fwd

input

Forward pass view 2

$$-y \ln x_1^{(m)} - (1-y) \ln \left(1 - x_1^{(m)}\right)$$

l

$x_1^{(m)}$

Forward pass view 1

$$x_i^{(k+1)} = \sigma\left(s_i^{(k+1)}\right)$$

$$s_i^{(k+1)} = \sum_{j=1}^{n_k} x_j^{(k)} w_{i,j}^{(k)}$$

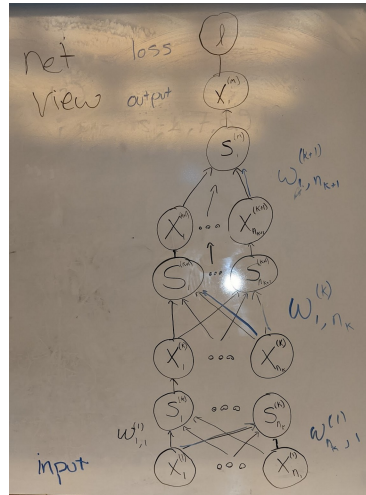$$w_{i,1}^{(k)} \quad \cdots \quad w_{i,n_k}^{(k)}$$

$$x_1^{(k)} \quad \cdots \quad x_n^{(k)}$$

Data flow view 3 with partial derivative on line

$y=1$ $y=0$

l $-y \ln x_1^{(m)} - (1-y) \ln \left(1 - x_1^{(m)}\right)$

$$-y\frac{1}{x_1^{(m)}} + (1-y)\frac{1}{1 - x_1^{(m)}} = \frac{\partial l}{\partial x_1^{(m)}}$$

$x_1^{(m)}$

Data flow view 1

$$\frac{\partial l}{\partial x_i^{(k+1)}}$$

$$x_i^{(k+1)} = \sigma\left(s_i^{(k+1)}\right) \quad x_i^{(k+1)}$$

$$\sigma(s_i^{(k+1)})(1 - \sigma(s_i^{(k+1)})) = \frac{\partial x_i^{(k+1)}}{\partial s_i^{(k+1)}}$$
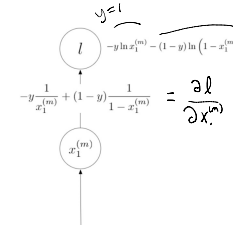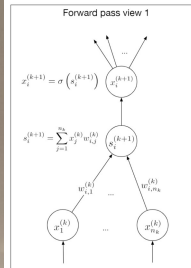
$$s_i^{(k+1)} = \mathbf{w}_i^{(k)\top} \mathbf{x}^{(k)} \quad s_i^{(k+1)}$$
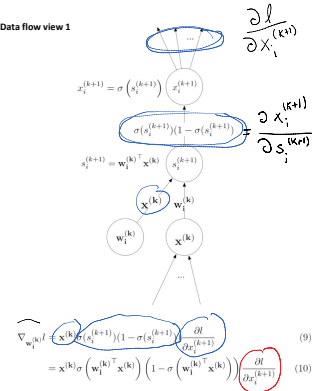
$$\mathbf{x}^{(k)} \quad \mathbf{w}_i^{(k)}$$

$$\mathbf{w}_i^{(k)} \quad \mathbf{x}^{(k)}$$

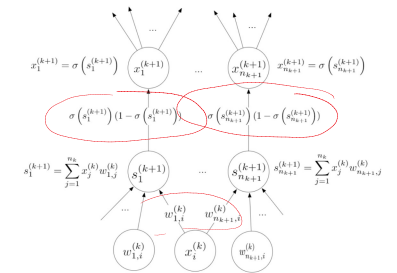$$\widehat{\nabla_{\mathbf{w}_i^{(k)}} l} = \mathbf{x}^{(k)} \sigma(s_i^{(k+1)})(1 - \sigma(s_i^{(k+1)})) \frac{\partial l}{\partial x_i^{(k+1)}} \tag{9}$$

$$= \mathbf{x}^{(k)} \sigma\left(\mathbf{w}_i^{(k)\top}\mathbf{x}^{(k)}\right) \left(1 - \sigma\left(\mathbf{w}_i^{(k)\top}\mathbf{x}^{(k)}\right)\right) \frac{\partial l}{\partial x_i^{(k+1)}} \tag{10}$$

Data flow view 2

$$x_1^{(k+1)} = \sigma\left(s_1^{(k+1)}\right) \quad x_1^{(k+1)} \quad \cdots \quad x_{n_{k+1}}^{(k+1)} \quad x_{n_{k+1}}^{(k+1)} = \sigma\left(s_{n_{k+1}}^{(k+1)}\right)$$

$$\sigma\left(s_1^{(k+1)}\right)\left(1 - \sigma\left(s_1^{(k+1)}\right)\right) \quad \sigma\left(s_{n_{k+1}}^{(k+1)}\right)\left(1 - \sigma\left(s_{n_{k+1}}^{(k+1)}\right)\right)$$

$$s_1^{(k+1)} = \sum_{j=1}^{n_k} x_j^{(k)} w_{1,j}^{(k)} \quad s_1^{(k+1)} \quad \cdots \quad s_{n_{k+1}}^{(k+1)} \quad s_{n_{k+1}}^{(k+1)} = \sum_{j=1}^{n_k} x_j^{(k)} w_{n_{k+1},j}^{(k)}$$

$$w_{1,i}^{(k)} \quad w_{n_{k+1},i}^{(k)}$$

$$w_{1,i}^{(k)} \quad x_i^{(k)} \quad w_{n_{k+1},i}^{(k)}$$

$$\frac{\partial l}{\partial x_i^{(k)}} = \sum_{j=1}^{n_{k+1}} w_{j,i}^{(k)} \sigma\left(s_j^{(k+1)}\right)\left(1 - \sigma\left(s_j^{(k+1)}\right)\right) \frac{\partial l}{\partial x_j^{(k+1)}}$$

need to consider all n_{k+1} nodes

$$= \sum_{j=1}^{n_{k+1}} w_{j,i}^{(k)} \sigma\left(\mathbf{w_j}^{(k)\top}\mathbf{x}^{(k)}\right)\left(1 - \sigma\left(\mathbf{w_j}^{(k)\top}\mathbf{x}^{(k)}\right)\right) \frac{\partial l}{\partial x_j^{(k+1)}} \tag{11}$$