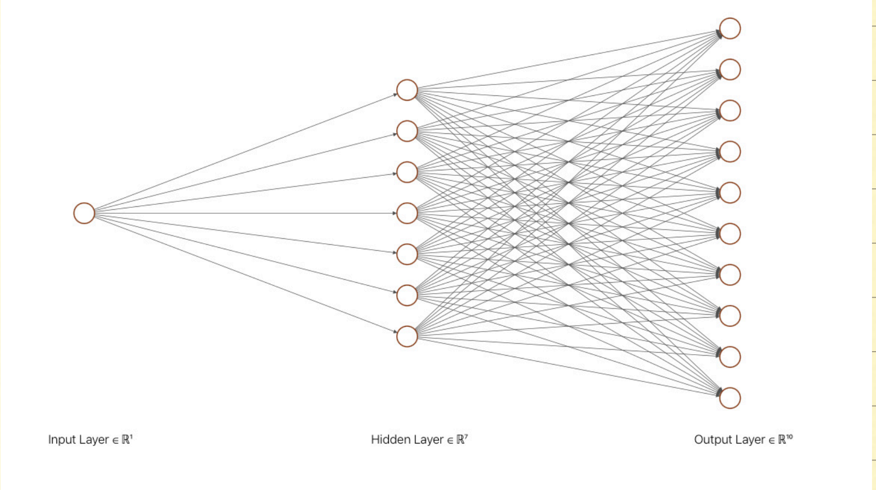

Gengcong Yan

1009903

Schematic network

1: $m=1$

$$E = -t \cdot \log y = -t \cdot \log(6(Wx)) = -t \cdot \log\left(\frac{1}{1+e^{-Wx}}\right)$$

$\therefore$

$$\frac{\partial E}{\partial t} = \log\left(\frac{1}{1+e^{-Wx}}\right) \qquad \frac{\partial E}{\partial x} = -t \cdot \frac{1}{6(Wx)} \cdot \frac{\partial 6(Wx)}{\partial x}$$

$$= -t \cdot (1+e^{-Wx}) \cdot W \frac{1}{1+e^{-Wx}}\left(1-\frac{1}{1+e^{-Wx}}\right)$$

$$= -tW \frac{e^{-Wx}}{1+e^{-Wx}}$$

2: There are only 2 layers in network, we can know:

$z^{(1)} = W^{(1)}x \qquad y^{(1)} = 6(z^{(1)})$,

$z^{(2)} = W^{(2)}y^{(1)} \qquad y^{(2)} = s(z^{(2)})$,

$y = y^{(2)}$

$$\frac{\partial E}{\partial z_i^{(2)}} = \sum_{j=1}^{n} \frac{\partial E_j}{\partial y_j^{(2)}} \cdot \frac{\partial y_j^{(2)}}{\partial z_i^{(2)}}$$

3: From (2) we can know: $\frac{\partial E}{\partial z^{(2)}} = (y^{(2)} - t)^T$

$\because z^{(2)} = W^{(2)} y^{(1)} \Rightarrow \frac{\partial z^{(2)}}{\partial y^{(1)}} = W^{(2)}$

$$\frac{\partial E}{\partial y^{(1)}} = \frac{\partial E}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial y^{(1)}} = (y^{(2)} - t)^T \cdot W^{(2)}$$

4: $$\frac{\partial E}{\partial W^{(2)}_{uv}} = \frac{\partial E}{\partial z^{(2)}} \cdot \frac{\partial z^{(2)}}{\partial W^{(2)}_{uv}} = (y^{(2)}_u - t_u)^T \cdot y^{(1)}_v$$

$$\therefore \frac{\partial E}{\partial W^{(2)}} = (y^{(2)} - t)\, y^{(1)T}$$

5:

$$\frac{\partial \sigma(z)}{\partial z} = \frac{-(-1)e^{z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} = \frac{1}{1+e^{-z}} \cdot \left(1 - \frac{1}{1+e^{-z}}\right)$$

$$= \sigma(z)\,(1 - \sigma(z))$$

$$\therefore \frac{\partial y^{(1)}}{\partial z^{(1)}} = \frac{\partial \sigma(z^{(1)})}{\partial z^{(1)}} = \sigma(z^{(1)}) \cdot (1 - \sigma(z^{(1)})) = y^{(1)} \cdot (1 - y^{(1)})$$

$$= \mathrm{diag}(y^{(1)} \cdot (1 - y^{(1)}))$$

6:

From the computation in (3) and (5), we can get:

$$\frac{\partial E}{\partial z^{(1)}} = \frac{\partial E}{\partial y^{(1)}} \cdot \frac{\partial y^{(1)}}{\partial z^{(1)}} = (y^{(2)} - t)^T\, W^{(2)}\, \mathrm{diag}(y^{(1)} .* (1 - y^{(1)}))$$

7:

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial Z^{(1)}} \cdot \frac{\partial Z^{(1)}}{\partial W^{(1)}} = \frac{\partial E}{\partial Z^{(1)}} \cdot X$$

$$\frac{\partial E}{\partial W_{uv}^{(1)}} = \frac{\partial E}{\partial Z_u^{(1)}} \cdot \frac{\partial Z_u^{(1)}}{\partial W_{uv}^{(1)}} = \frac{\partial E}{\partial Z_u^{(1)}} \cdot \frac{\partial (\sum_i W_{ui} X_i)}{\partial W_{uv}^{(1)}} = \frac{\partial E}{\partial Z_u^{(1)}} \cdot X_v$$