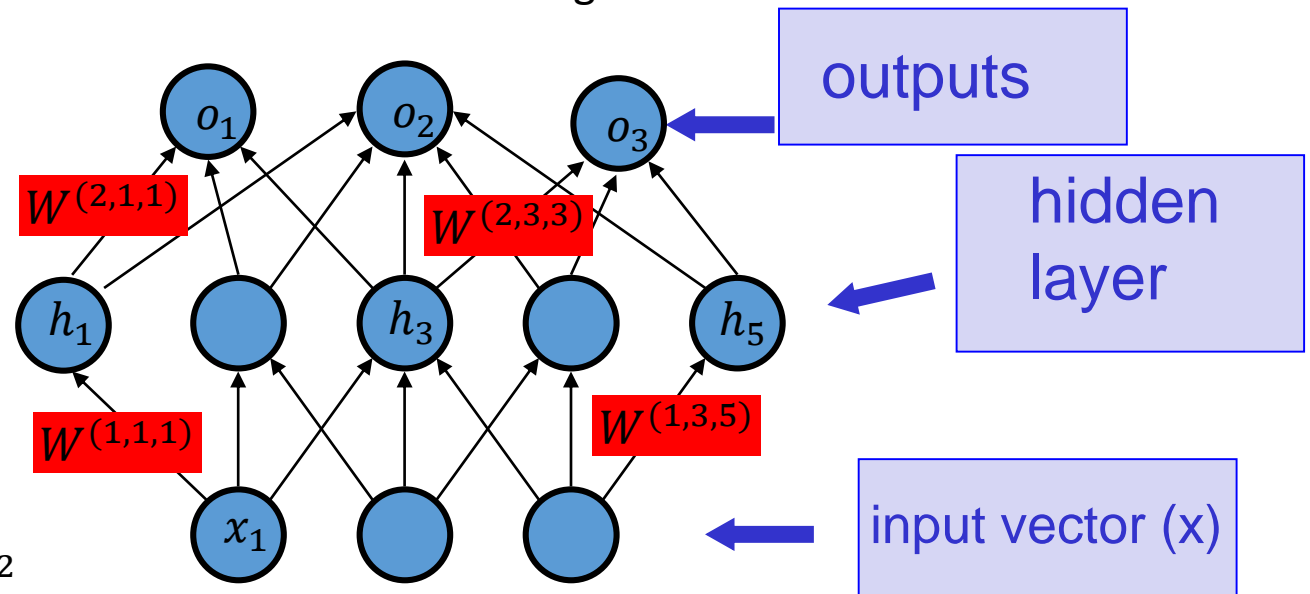


# Multilayer Neural Network for Classification (to be modified later)

$o_i$  is large if the probability that the correct class is  $i$  is high



A possible cost function:

$$\sum_{i=1}^m (o^{(i)} - y^{(i)})^2$$

$y^{(i)}$ 's encoded using one-hot encoding

# Aside: Vectorizing the ANN

- $o_i = g\left(\sum_j W^{(2,j,i)} h_j + b^{(2,j)}\right)$

- So  $o = g\left(\left(W^{(2)}\right)^T h + b^{(2)}\right)$

$g$ : activation function  
e.g.  $g(t) = \frac{1}{1+\exp(-t)}$

- Similarly,  $h = g\left(\left(W^{(1)}\right)^T x + b^{(1)}\right)$