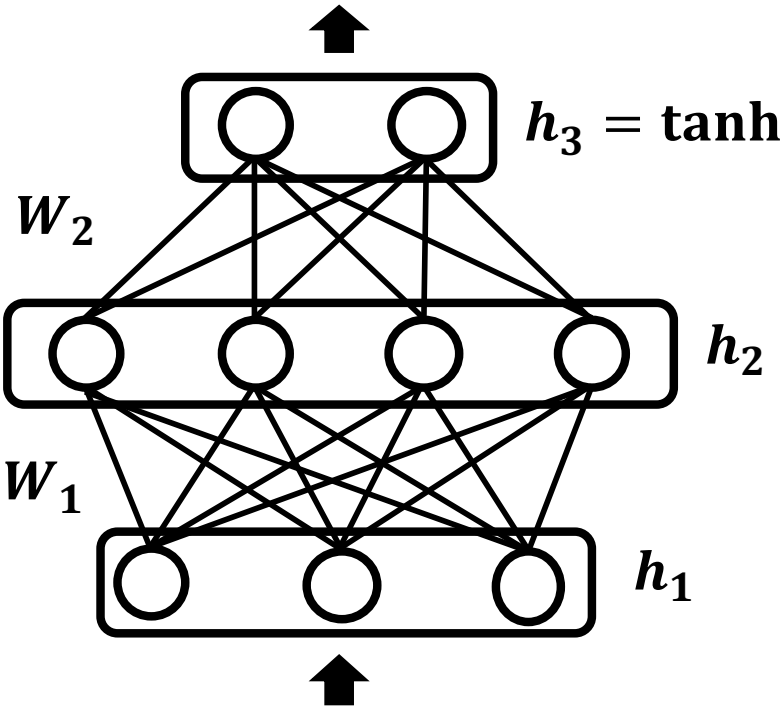
Deep Learning

31

$f(x) = p(female) =$ 0.9



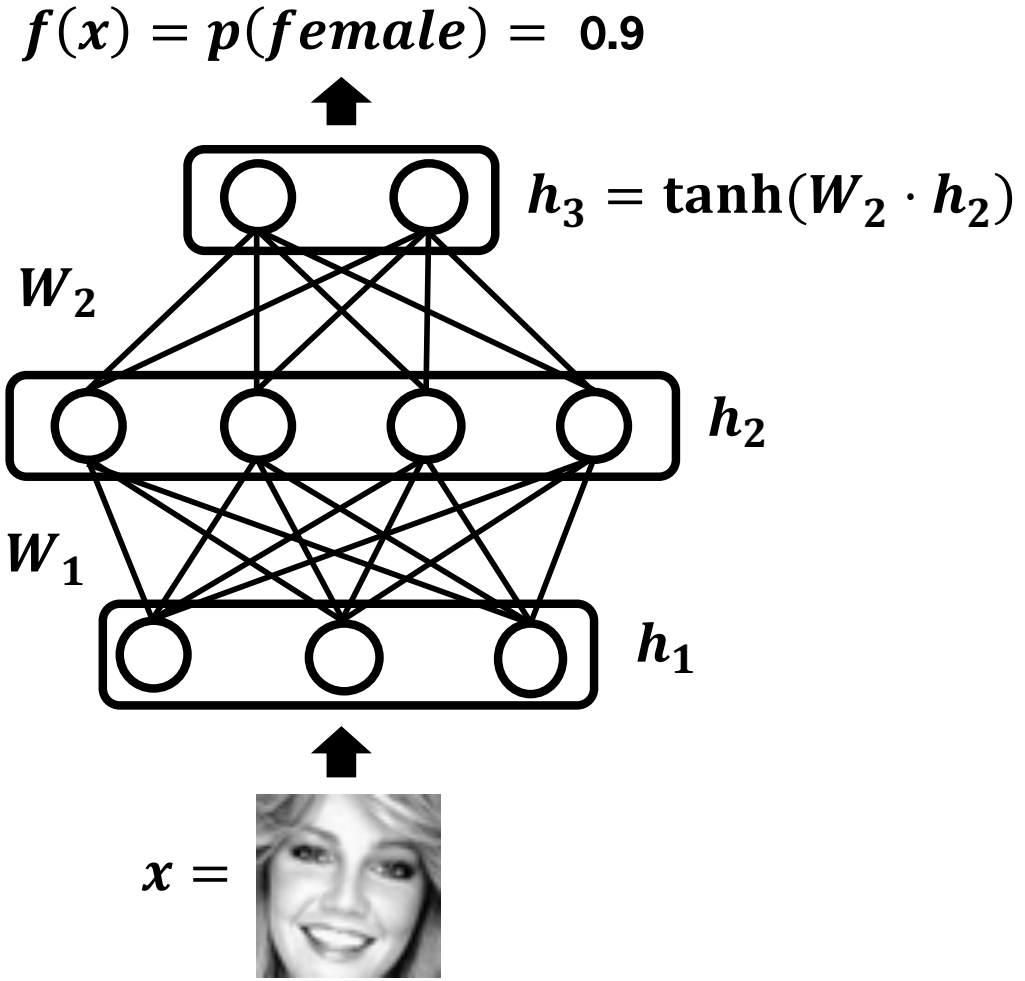
$x =$ 

- Map input x to layers of hidden representations h, then to output y
- $h_{l+1} = a(W_l \cdot h_l)$ with parameter W_l
- Train model to minimizes loss:

$$W = \text{argmin}_W L(f(x), y)$$

- Gradient descent on parameters:
  - Each iteration train on a batch
  - Update W based on gradient of L

Deep Learning



- Map input x to layers of hidden representations h, then to output y
- $h_{l+1} = a(W_l \cdot h_l)$ with parameter W_l
- Train model to minimizes loss:

$$W = \text{argmin}_W L(f(x), y)$$

- Gradient descent on parameters:
  - Each iteration train on a batch
  - Update W based on gradient of L

Leakage From Model Updates