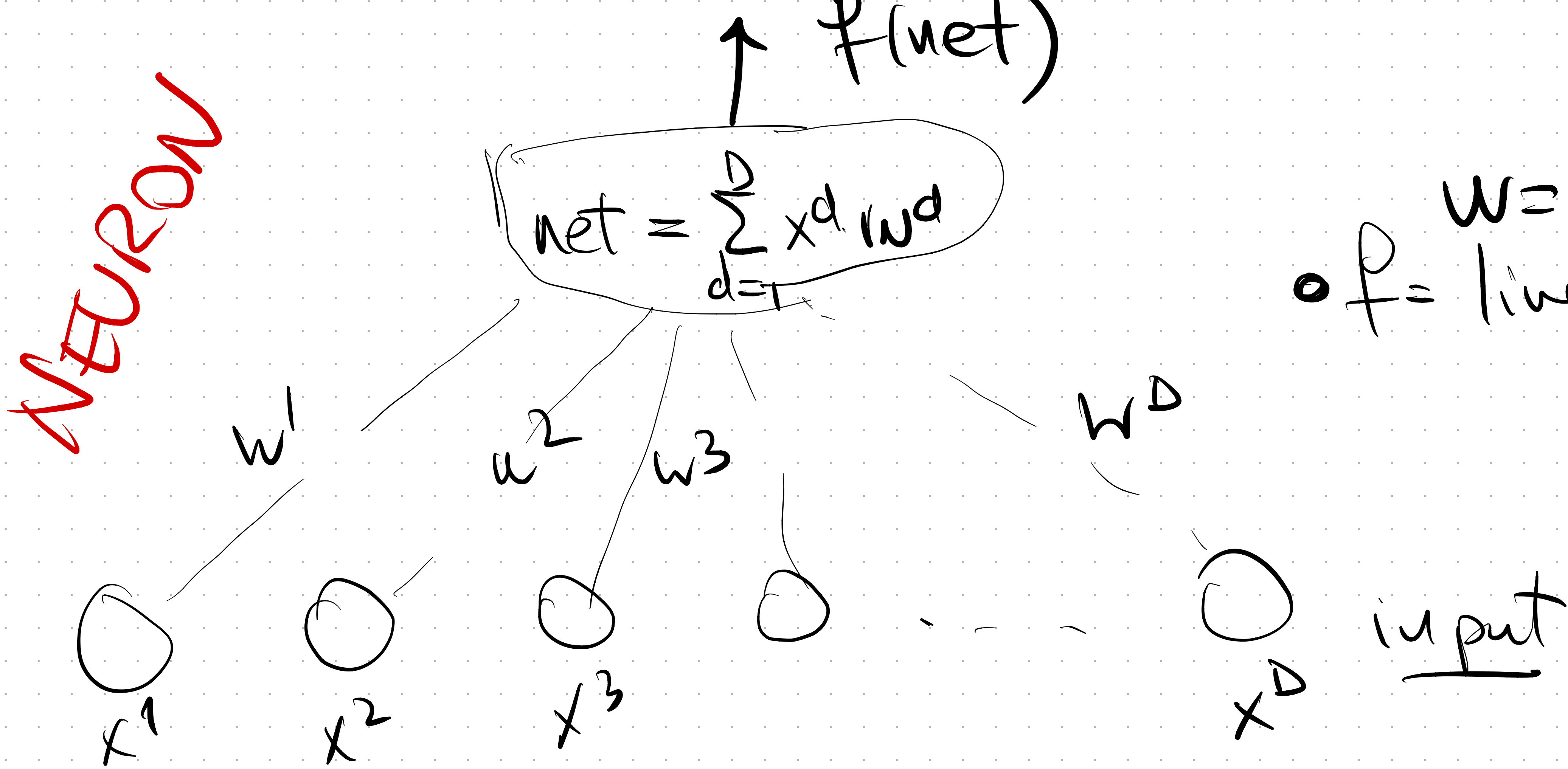
5/30: Neural Networks (part 1: very basics) $\boxed{\text{Later}}$ p2: adv. NN
p3: recurrent/atten

Recap: Linear Model: Lin Reg, Logistic Reg, Perceptron

net = linear
dot product
$\langle x w \rangle$

Error = OBJ(w)



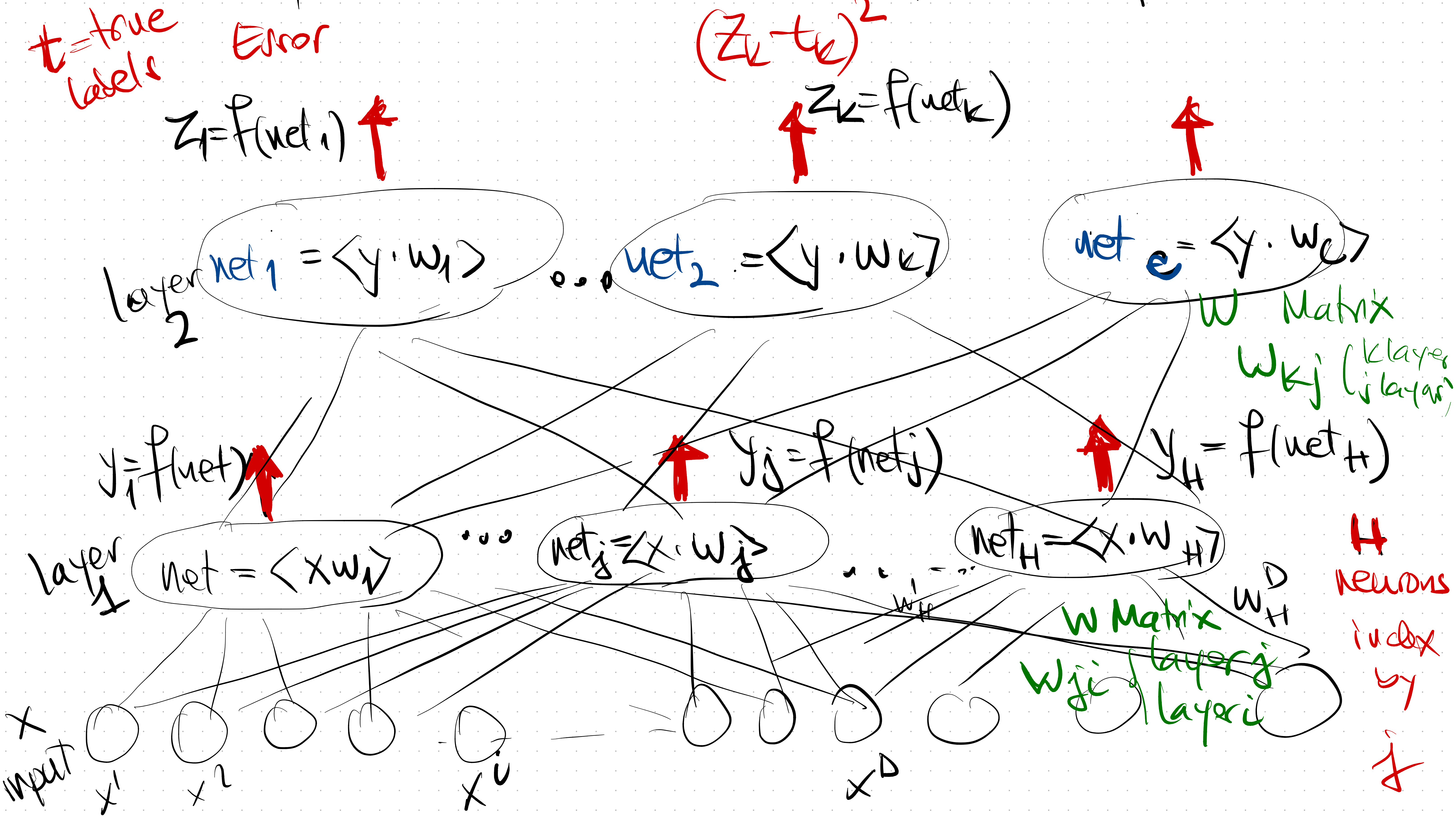
f = activator
w = model weights

- f = linear / identity $\Rightarrow$ output is linear
  - Lin Reg, - Percep
- f non linear
  ex f = sigmoid/logistic

GD: $\frac{\partial \text{Error}}{\partial w} \Longrightarrow$ update weights to improve Error $\Rightarrow$ output $\neq$ linear

"back-propagate" the error to weights

Stack of neurons : output of layer = input to next layer.



GD update: Error $\Rightarrow$ update weights

- start at the top (Error – top layer) $\frac{\partial Error}{\partial w} \Rightarrow$ update W top layer
- mor to second layer (from top) backpropagate Error $\Rightarrow$ update W on second layer

Simple Network (input – hidden layer 1 – output layer – error)

OBJ $J = \frac{1}{2} \sum_k (t_k - z_k)^2 = \sum_k (t_k - f(net_k))^2 = \sum_k (t_k - f(y \cdot w_k))^2$

top layer

- update weights at top layer w_{kj}

$$\frac{\partial J}{\partial w_{kj}} = \frac{\partial J}{\partial net_k} \cdot \frac{\partial net_k}{\partial w_{kj}} = \frac{\partial J}{\partial z_k} \cdot \frac{\partial z_k}{\partial net_k} \cdot \frac{\partial (y \cdot w_k)}{\partial w_{kj}}$$

$$= (z_k - t_k) \cdot \boxed{\frac{\partial f}{\partial net_k}} \cdot y_j$$

$f'(net_k)$

f = non linear activator

GD update

learn rate

$$w_{kj(new)} = w_{kj} - \lambda (z_k - t_k) f'(net_k) \cdot y_j$$

- update the bottom layer w_{ji}

$$\frac{\partial J}{\partial w_{ji}} = \frac{\partial J}{\partial y_j} \cdot \frac{\partial y_j}{\partial net_j} \cdot \boxed{\frac{\partial net_j}{\partial w_{ji}}} \qquad \frac{\partial (x \cdot w_j)}{\partial w_{ji}}$$

$$\frac{\partial J}{\partial y_j} \to \frac{\partial \frac{1}{2}\left[\sum_k (t_k - z_k)^2\right]}{\partial y_j} \qquad \frac{\partial y_j}{\partial net_j} \to f'(net_j) \qquad \frac{\partial net_j}{\partial w_{ji}} \to x^i$$

$$\underbrace{f'(net_j) \quad x^i} \to f'(net_j) \cdot x^i$$

$$-\sum_k (t_k - z_k) \boxed{\frac{\partial z_k}{\partial net_k}} \cdot \boxed{\frac{\partial net_k}{\partial y_j}} \qquad \frac{\partial z_k}{\partial net_k} = f'(net_k) \qquad \frac{\partial (y, w_k)}{\partial y_j} = w^k_j$$

$$-\left[\sum_k (t_k - z_k) f'(net_k) \cdot w^k\right] \cdot f'(net_j) \cdot x^i$$

GD update

$$w_{ji(new)} = w_{ji} + \lambda \left[\sum_k (t_k - Z_k) f'(net_k) w_{kj}\right] \cdot f'(net_j) \cdot x^i$$