Network with 1 Hidden Layer

Input layer: $x_1, \ldots, x_d$; $x_{d+1} = 1$

Hidden units: $h_1, \ldots, h_m$; $h_{m+1} = 1$

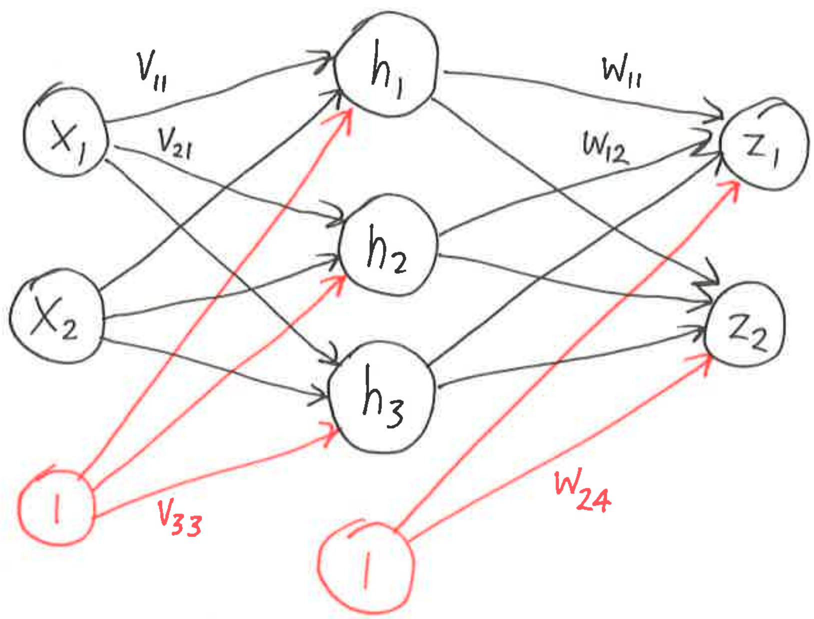
Output layer: $z_1, \ldots, z_k$ [We might have more than one output so we can build multiple classifiers that share hidden units.]

Layer 1 weights: $m \times (d+1)$ matrix V $\quad V_i$ is row i

Layer 2 weights: $k \times (m+1)$ matrix W $\quad W_i$ is row i

Recall $s(\gamma) = \frac{1}{1+e^{-\gamma}}$ other nonlinear fns can be used

For vector v, $s(v) = \begin{bmatrix} s(v_1) \\ s(v_2) \\ \vdots \end{bmatrix}$ [We apply s to a vector component by component]

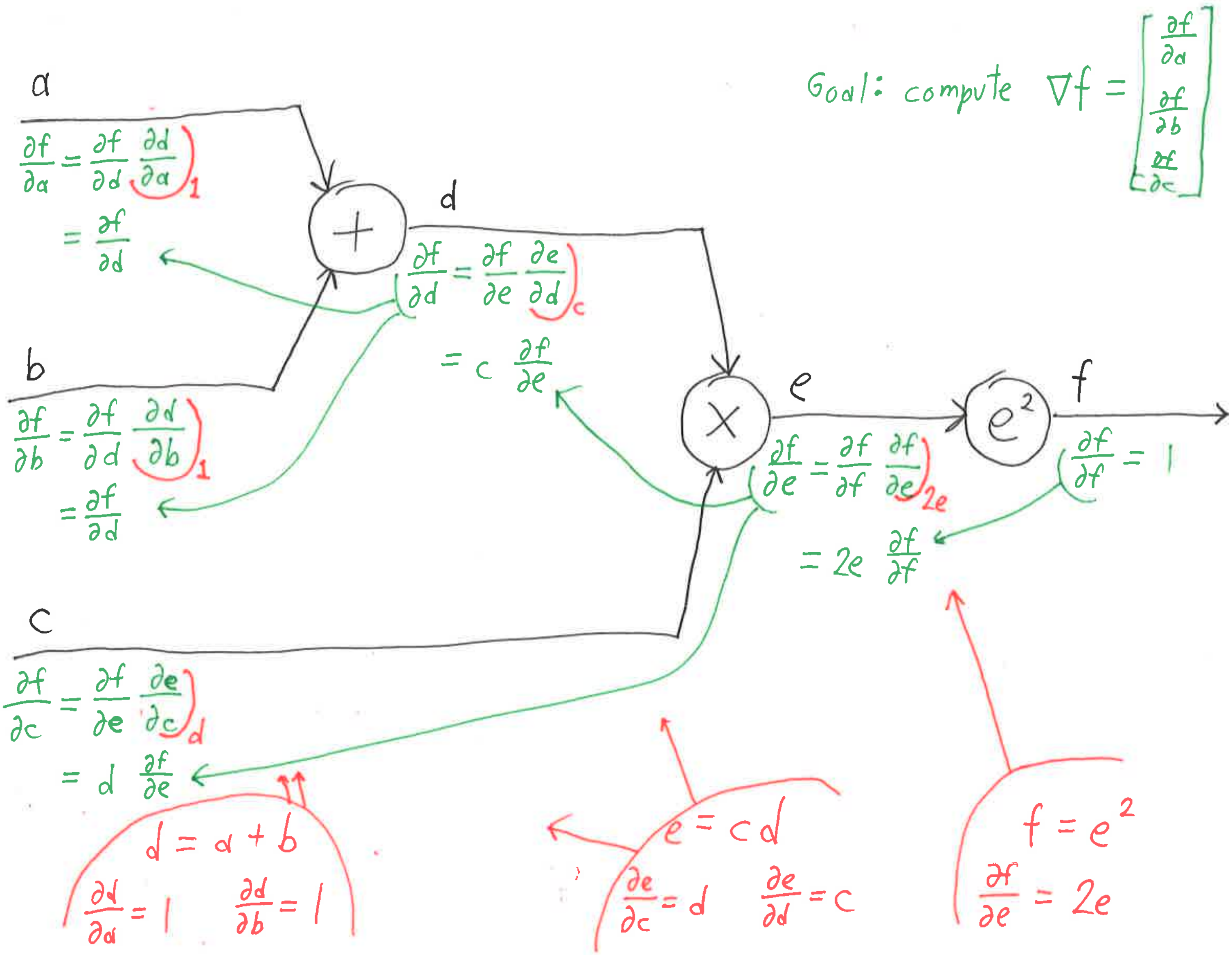


$h_i = s\left(\sum_{j=1}^{3} V_{ij}\, x_j\right)$ In short, $h = s(Vx)$

$z = s(Wh) = s(W\, s_1(Vx))$

add a 1 to end of vector

Computing Gradients for Arithmetic Expressions



Each value z ~~has~~ gives partial derivative of the form

$$\frac{\partial f}{\partial z} = \underbrace{\frac{\partial f}{\partial n}}_{\text{computed during backward pass after forward pass}} \underbrace{\frac{\partial n}{\partial z}}_{\text{computed during forward pass}}$$

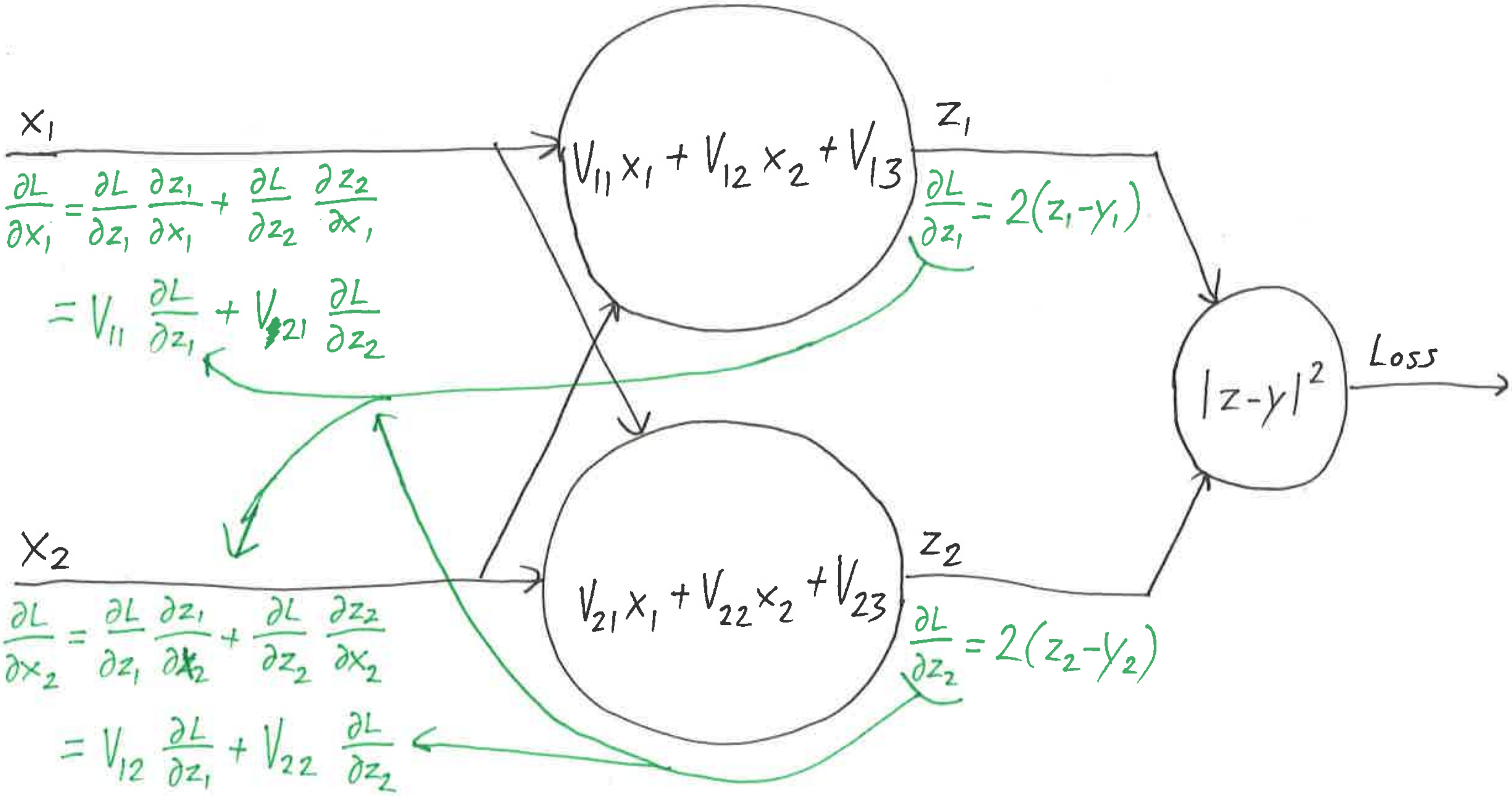
where z is input to n.

computed during forward pass

computed during backward pass <u>after</u> forward pass

"backpropagation"

[What if a unit's output goes to more than one unit?]



$$\frac{\partial L}{\partial z_1} = 2(z_1 - y_1)$$

$$\frac{\partial L}{\partial z_2} = 2(z_2 - y_2)$$

$$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

$$= V_{11}\frac{\partial L}{\partial z_1} + V_{21}\frac{\partial L}{\partial z_2}$$

$$\frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$

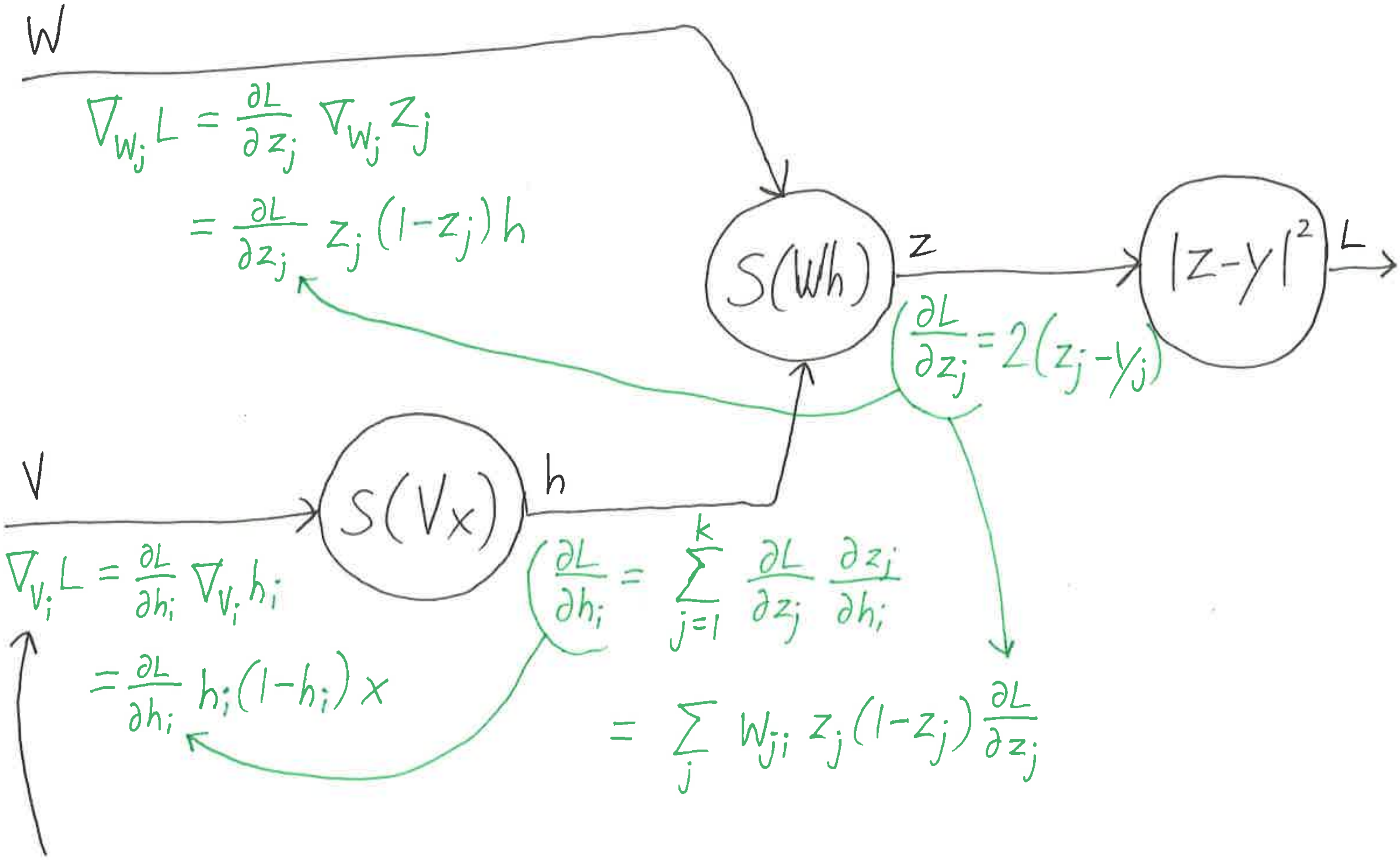
$$= V_{12}\frac{\partial L}{\partial z_1} + V_{22}\frac{\partial L}{\partial z_2}$$

The Backpropagation Alg

Recall $s'(\gamma) = s(\gamma)(1-s(\gamma))$

$h_i = s(V_i \cdot x)$, so $\nabla_{V_i} h_i = s'(V_i \cdot x)\, x$
$\quad\quad = h_i(1-h_i)\, x$

$\nabla_{W_j} z_j = s'(W_j \cdot h)\, h$
$\quad\quad = z_j(1-z_j)\, h$



$$\nabla_{W_j} L = \frac{\partial L}{\partial z_j} \nabla_{W_j} z_j = \frac{\partial L}{\partial z_j} z_j(1-z_j)\, h$$

$$\frac{\partial L}{\partial z_j} = 2(z_j - y_j)$$

$$\nabla_{V_i} L = \frac{\partial L}{\partial h_i} \nabla_{V_i} h_i = \frac{\partial L}{\partial h_i} h_i(1-h_i)\, x$$

$$\frac{\partial L}{\partial h_i} = \sum_{j=1}^{k} \frac{\partial L}{\partial z_j} \frac{\partial z_j}{\partial h_i} = \sum_j W_{ji}\, z_j(1-z_j) \frac{\partial L}{\partial z_j}$$

Compute $\nabla_V L$, $\nabla_W L$ one row at a time.