

Backpropagation Pseudo-code

→ this will be different for diff f' 's.

$$\nabla_{a_L} \mathcal{L}(\theta) = - (y - f(m)) = \hat{y} - y \text{ (softmax)}$$

For $k=L$ to $k=1$ do;

$$\nabla_{w_k} \mathcal{L}(\theta) = \nabla_{a_k} \mathcal{L}(\theta) \cdot h_{k-1}^T$$

$$\nabla_{b_k} \mathcal{L}(\theta) = \nabla_{a_k} \mathcal{L}(\theta)$$

$$\nabla_{h_{k-1}} \mathcal{L}(\theta) = w_k^T (\nabla_{a_k} \mathcal{L}(\theta))$$

$$\nabla_{a_{k-1}} \mathcal{L}(\theta) = \nabla_{h_{k-1}} \mathcal{L}(\theta) \odot \left(\dots g'(a_{k-1,j}) \dots \right)$$

Adamvrd
prod.

