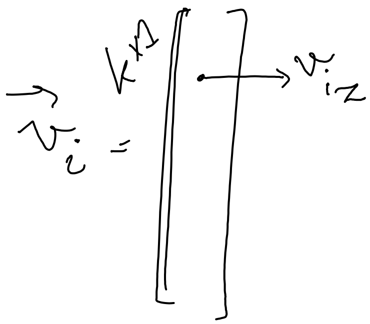
Lecture 10/17: Backpropagation

Thursday, October 17, 2019 2:48 PM

Vectorized



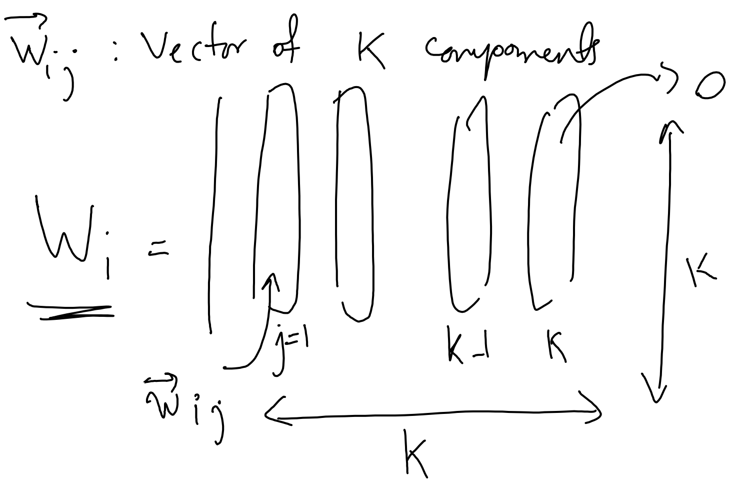
$z \in [k]$

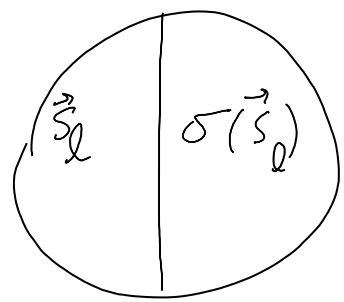
values in layer i

$$\begin{pmatrix} \sigma(\vec{v}_i \cdot \vec{w}_{i1}) \\ \vdots \\ \sigma(\vec{v}_i, w_{i,k-1}) \\ 1 \end{pmatrix} = \sigma \begin{pmatrix} \vec{v}_i \cdot \vec{w}_{i1} \\ \vdots \\ \vec{v}_i \cdot \vec{w}_{ik} \end{pmatrix}$$

$$= \sigma\left(W_i^T \vec{v}_i\right)$$

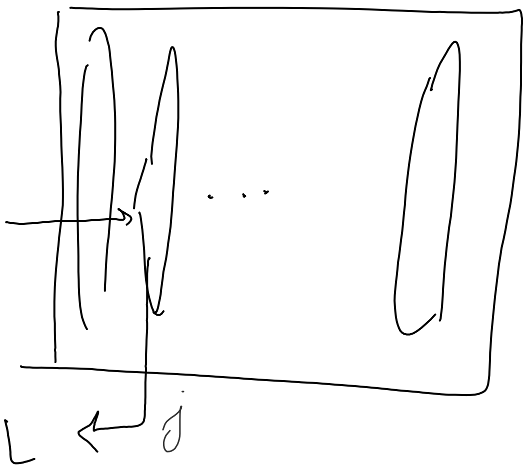
$\vec{w}_{i,j}$: Vector of K components





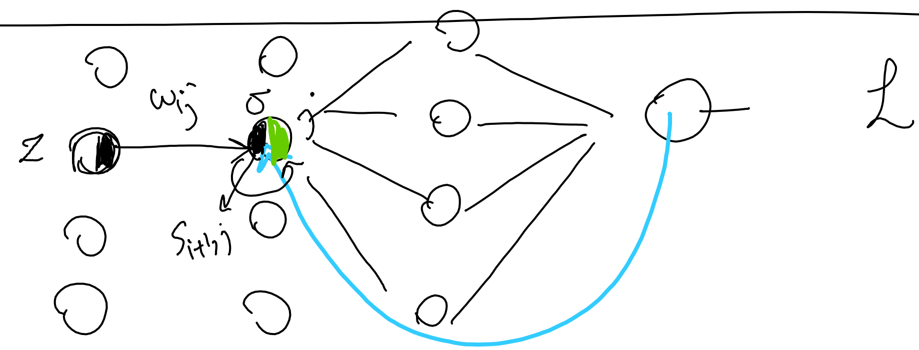
$\frac{\partial L}{\partial w_1}$

z



$\frac{\partial L}{\partial w_{ij}^{(2)}}$

$$\frac{\partial L}{\partial w_{ij}(2)} =$$



$$\frac{\partial L}{\partial v_{i+1,j}} \cdot \underbrace{\frac{\partial v_{i+1,j}}{\partial S_{i+1,j}}}_{\sigma'(S_{i+1,j})} \cdot \underbrace{\frac{\partial S_{i+1,j}}{\partial w_{ij}(2)}}_{V_{i2}}$$

$V_{.} \cdot S_{+1,j})$

$$S_{i+1,j} = \vec{V}_i \circ \vec{W}_{ij}$$

$\uparrow\downarrow$ $\frac{\partial L}{w_{ij}(2)}$

$\downarrow$

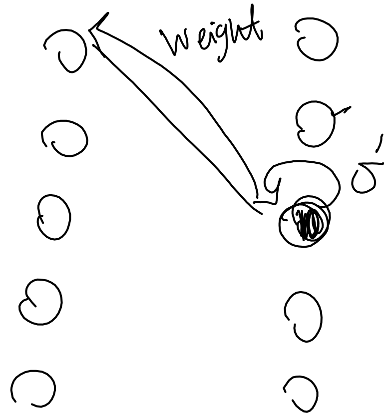
$$\frac{\partial L}{\partial v_{ij}}$$

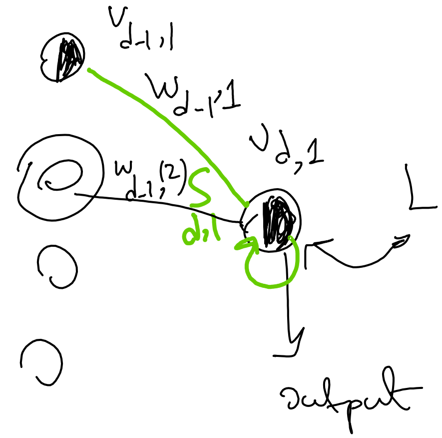
forward prediction

$$\boxed{L(y',y) = \frac{1}{2}(y'-y)^2}$$

$$\frac{\partial L}{\partial v_{d,1}} = \frac{\partial}{\partial v_{d,1}}\left(\frac{1}{2}(v_{d,1}-y)^2\right)$$

$$= v_{d,1} - y$$

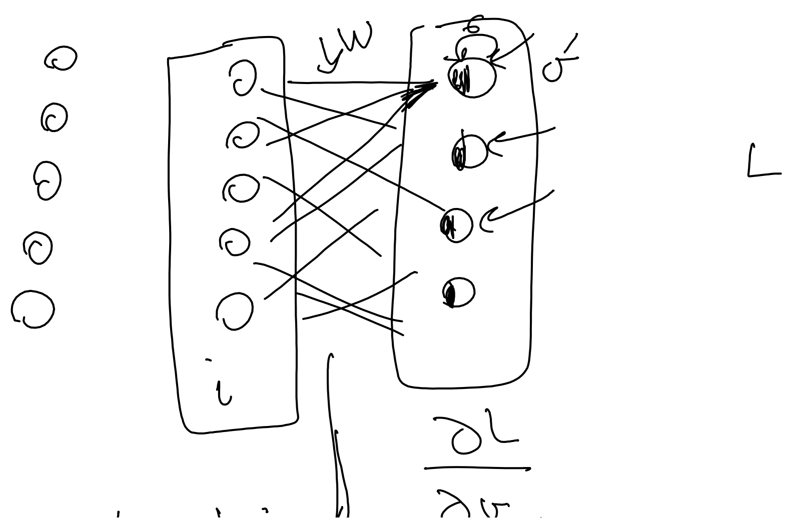




$$\frac{\partial L}{\partial v_{d-1,1}} = \underbrace{\frac{\partial L}{\partial v_{d,1}}}_{\checkmark} \cdot \underbrace{\frac{\partial v_{d,1}}{\partial s_{d,1}}}_{\sigma'(s_{d,1})} \cdot \underbrace{\frac{\partial s_{d,1}}{\partial v_{d-1,1}}}_{\frac{w_{d-1,1}(1)}{w_{d-1,1}(2)}\ \text{or}}$$

$$\frac{\partial L}{\partial \vec{v}_i} = \underbrace{\frac{\partial L}{\partial \vec{v}_{i+1}}}_{\frac{\partial L}{\partial \vec{v}_{i+1}}} \cdot \underbrace{\frac{\partial \vec{v}_{i+1}}{\partial \vec{s}_{i+1,1}}}_{\sigma'(\vec{s}_{i+1})} \cdot \underbrace{\frac{\partial \vec{s}_{i+1,1}}{\partial \vec{v}_i}}_{W_i}$$

layer



$k \times k$: W_i ... $_{i+1}$

$$\frac{\partial L}{\partial W_i} = \frac{\partial L}{\partial \vec{v}_{i+1}} \cdot \underbrace{\frac{\partial \vec{v}_{i+1}}{\partial \vec{s}_{i+1}}}_{\sigma'(\vec{s}_{i+1})} \cdot \underbrace{\frac{\partial \vec{s}_{i+1}}{\partial W_i}}_{V_i}$$

$k \times 1$

At a high level:

keep track of $\quad \vec{\delta}_d = \frac{\partial L}{\partial \vec{s}_d} \qquad \vec{\delta}_{d-1} = \frac{\partial L}{\partial s_d} \cdot \frac{\partial s_d}{\partial v_{d-1}} \cdot \frac{\partial v_{d-1}}{\partial s_{d-1}}$

At layer ℓ: $\boxed{\vec{\delta}_\ell = W_\ell \left(\sigma'(\vec{s}_\ell) \odot \vec{\delta}_{\ell+1} \right)}$

$$W_\ell \leftarrow W_\ell - \eta \underbrace{\frac{\partial L}{\partial W_\ell}}_{} = W - \underbrace{\frac{\partial L}{\partial s_{\ell+1}}}_{\vec{\delta}_{\ell+1}} \cdot \underbrace{\frac{\partial s_{\ell+1}}{\partial W_\ell}}_{\vec{v}_\ell}$$