ELCmvD

ELCmvD

Lecture - 20

~~Fitting the~~ Gradient of E_{in}

Given weights W,

$$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} e_n = \frac{1}{N}\sum_{n=1}^{N} e\left(h(x_n; w), y_n\right)$$

the gradient of $E_{in}(w)$ is:

$$\nabla$$

* **Backpropagation** :-

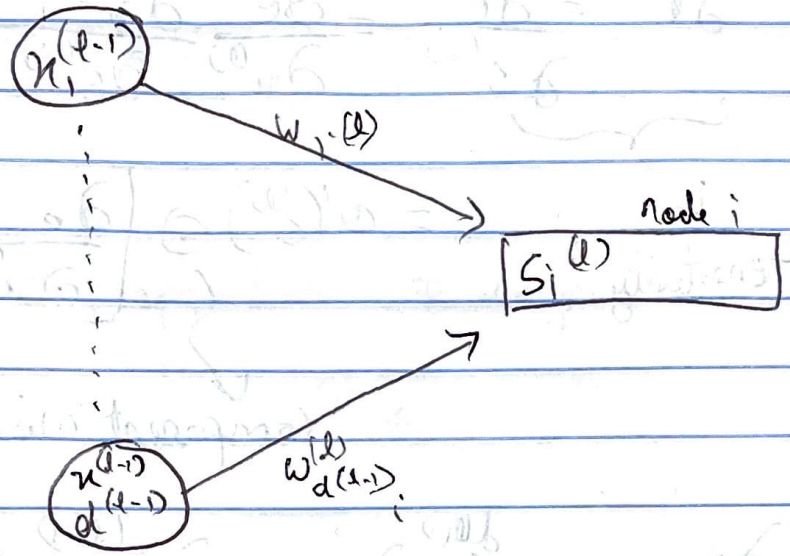
computing the ~~de~~ gradient: Using the chain rule.

Node i only affects the output $h(x)$ through its inputs $s_i^{(l)}$.

$$\frac{\partial e}{\partial w_i^{(l)}} = \frac{\partial s_i^{(l)}}{\partial w_i^{(l)}} \cdot \frac{\partial e}{\partial s_i^{(l)}}$$

- $\frac{\partial e}{\partial w_i^{(l)}}$: How error changes w.r.t $w_i^{(l)}$
- $\frac{\partial s_i^{(l)}}{\partial w_i^{(l)}}$: How input signal changes w.r.t $w_i^{(l)}$
- $\frac{\partial e}{\partial s_i^{(l)}}$: sensitivity.
- (chain rule).

$$s_i^{(l)} = (w_i^{(l)})^T x^{(l-1)} \Rightarrow \frac{\partial s_i^{(l)}}{\partial w_i^{(l)}} = x^{(l-1)}$$



if $l > 1$

$$x^{(l-1)} = \tanh\left(\overleftarrow{s_i^{(l-1)}} \; T \; w^{(l-1)} \; x^{(l-2)} \right)$$

$$d^{(l-1)} \times d^{(l-2)} \quad (1 + d^{(l-3)})$$

$$x^{(l-1)} = \begin{bmatrix} 0 \\ s^{(l-1)} \end{bmatrix} \rightarrow \tanh.$$

$$\frac{\partial e}{\partial w_i^{(l)}} = x^{(l-1)} \cdot (\$ s^{(l)})^T$$

Sensitivity

* **Sensitivity** : Using the chain rule Again

$$\delta^{(l)} = \frac{\partial e}{\partial s^{(l)}} = \frac{\partial e}{\partial x^{(l)}} \frac{\partial x^{(l)}}{\partial s^{(l)}}$$

$\delta^{(l)}$ ↙ Sensitivity

$$= \theta'(s^{(l)}) \odot \left[\frac{\partial e}{\partial x^{(l)}}\right]_{j}^{d^{(l)}}$$

↓ Component wise multiplication

$$\frac{\partial e}{\partial x^{(l)}} = \sum_{k=1}^{d^{(l+1)}} \frac{\partial e}{\partial s_k^{(l+1)}} \cdot \frac{\partial s_k^{(l+1)}}{\partial x^{(l)}}$$

$\frac{\partial e}{\partial s_k^{(l+1)}} \rightarrow \delta^{(l+1)}$; $\frac{\partial s_k^{(l+1)}}{\partial x^{(l)}} \rightarrow w_k^{(l+1)}$

* We can now compute,

$\delta^{(l)} = ?$

$$\delta^{(l)} \leftarrow \delta^{(l+1)} \leftarrow \dots \delta^{L-1} \leftarrow \delta^{L}$$

x^{L} is just the tanh applied to $\frac{\partial x^{(L)}}{\partial s^{(L)}}$

$$\frac{\partial (x^{(L)} - y)^2}{\partial s^{(L)}}$$

$$= 2(x^{(L)} - y) \frac{\partial x^{(L)}}{\partial s^{(L)}}$$

* Backpropagation to compute sensitivities.

$$\delta^{(1)} \leftarrow \ldots \leftarrow \delta^{(L-2)} \leftarrow \delta^{(L-1)} \leftarrow \delta^{(L)}.$$

Using backpropagation we will calculate the sensitivities and then we will use these sensitives to calculate gradient descent.

We will use Forward propagation for outputs
Backward propagation for sensitivity