CW 2

Tuesday, February 16, 2021 9:23 AM

1) Softmax with back propagation

i : node in input
j : node in hidden layer
k : node in output

} from input to hidden layer: ω_{ji}

Activation of hidden unit : $a_j = f(z_j)$, $z_j = \sum_{i=0}^{I} \omega_{ji} x_i$

Activation of output unit : $\hat{y}_k = f(z_k)$, f = output unit activation function

From CW1:

① $\omega_{kj} := \omega_{kj} - \alpha \frac{\partial C}{\partial \omega_{kj}} = \omega_{kj} - \alpha \delta_k a_j$

② $\omega_{ji} := \omega_{ji} - \alpha \frac{\partial C}{\partial \omega_{ji}}$, ③ $\delta_j := \frac{\partial C}{\partial z_j}$

a) Show that I) $\omega_{ji} = \omega_{ji} - \alpha \delta_j x_i$ and that

II) $\delta_j = f'(z_j) \sum_k \omega_{kj} \delta_k$

I) Hint ②: $\delta_k = \frac{\partial C}{\partial z_k} = -(y_k - \hat{y}_k)$

Hint ① : Use chain rule on $\alpha \frac{\partial C}{\partial \omega_{ji}}$

i) Computing $\frac{\partial C}{\partial \omega_{ji}}$:

$$\frac{\partial C}{\partial \omega_{ji}} = \sum_k \underbrace{\frac{\partial C}{\partial z_k}}_{\mathrm{I}} \underbrace{\frac{\partial z_k}{\partial a_j}}_{\mathrm{II})} \underbrace{\frac{\partial a_j}{\partial z_j}}_{\mathrm{III})} \underbrace{\frac{\partial z_j}{\partial \omega_{ji}}}_{\mathrm{IV})}$$

I) $\frac{\partial C}{\partial z_k} = \delta_k$ II) $\frac{\partial z_k}{\partial a_j} = \frac{\partial}{\partial a_j} \sum_j \omega_{kj} a_j = \sum_j \omega_{kj}$

III) $\frac{\partial a_j}{\partial z_j} = \frac{\partial f(z_j)}{\partial z_j} = f'(z_j)$

IV) $\frac{\partial z_j}{\partial \omega_{ji}} = \frac{\partial}{\partial \omega_{ji}} \sum_i \omega_{ji} x_i = \sum_i x_i$

$\Rightarrow \partial C$ …

$$\Rightarrow \frac{\partial C}{\partial \omega_{ji}} = \delta_k \sum_j \omega_{kj} f'(z_j)\, x_i \quad , \text{ where } \quad \underline{\underline{\delta_j = \delta_k \sum_j \omega_{kj} f'(z_j)}}$$

$$= \delta_j x_i$$

$$\Rightarrow \underline{\underline{\omega_{ji} = \omega_{ji} - \alpha \overbrace{\delta_j x_i}^{\frac{\partial C}{\partial \omega_{ji}}}}}$$

1b) Vectorize computation: Update ω_{ji} and ω_{kj} at the same time

I) $\omega_{ji} = \omega_{ji} - \alpha \delta_j x_i = \omega_{ji} - \alpha \delta_k \sum_k \omega_{kj} f'(z_j)$

II) $\omega_{kj} = \omega_{kj} - \alpha \delta_k a_j$

Idea: $\vec{W}_{kj} = \vec{W}_{kj} - \alpha \vec{\delta}_k \vec{a}_j^T$

$$\vec{W}_{ji} = \vec{W}_{ji} - \alpha \vec{\delta}_j \vec{x}_i^T = \vec{W}_{ji} - \alpha \underbrace{\left(\sum_k \vec{\delta}_k \vec{\omega}_{kj}\right) f'(\vec{z}_j)}_{\vec{\delta}_j} \vec{x}_i^T$$

$$= \vec{W}_{ji} - \alpha \vec{F}'(\vec{z}_j) \sum_k \overrightarrow{\delta_k \omega_{kj}}\, \vec{x}_i^T$$

$$\underline{\underline{= \vec{W}_{ji} - \alpha \vec{F}'\left(\sum_{i=0}^{I} \vec{\omega}_{ji}^T \vec{x}_i\right) \sum_k \left(\vec{\delta}_k \vec{\omega}_{kj}^T\right) \vec{x}_i^T}}$$