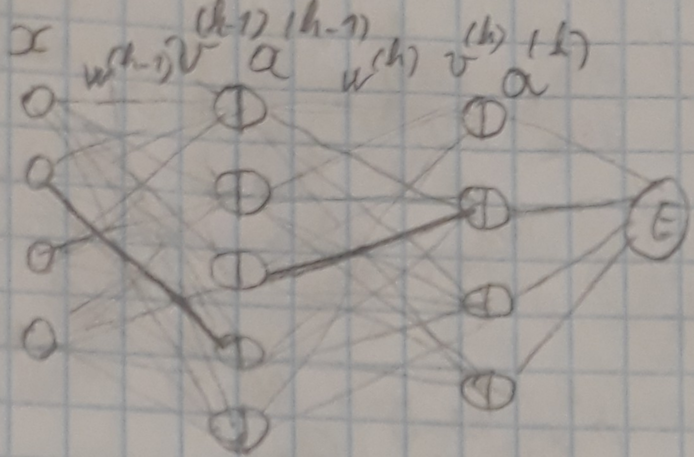


$W \in \mathbb{R}^{(n\_outputs,\ n\_inputs)}$

расширеные данные единицами

$$u_k = \sum_{j=1}^{J} w_{k,j} x_j \qquad v_k = u_k + b_k = \sum_{j=0}^{J} w_{k,j} x_j$$

$$a_k = \varphi(u_k + b_k)$$

$$E(\varphi^{(h)}(w^{(h)} a^{(h-1)})) = E(\varphi^{(h)}(w^{(h)} \varphi^{(h-1)}(w^{(h-1)} a^{(h-2)})))$$

$$a^{(0)} = x$$

$$\frac{\partial E}{\partial w_{i,j}^{(h)}} = \frac{\partial E}{\partial a_i^{(h)}} \frac{\partial a_i^{(h)}}{\partial w_{i,j}^{(h)}} = \frac{\partial E}{\partial a_i^{(h)}} \frac{\partial a_i^{(h)}}{\partial v_i^{(h)}} \frac{\partial v_i^{(h)}}{\partial w_{i,j}^{(h)}} = \frac{\partial E}{\partial v_i^{(h)}} a_j^{(h-1)} = \delta_i^{(h)} a_j^{(h-1)}$$

$$\frac{\partial E}{\partial w_{i,j}^{(h-1)}} = \sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial w_{ij}^{(h-1)}} = \sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial v_k^{(h)}} \frac{\partial v_k^{(h)}}{\partial w_{ij}^{(h-1)}}$$

$$= \sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial v_k^{(h)}} \frac{\partial v_k^{(h)}}{\partial a_i^{(h-1)}} \frac{\partial a_i^{(h-1)}}{\partial v_i^{(h-1)}} \frac{\partial v_i^{(h-1)}}{\partial w_{ij}^{(h-2)}}$$

$$= \frac{\partial E}{\partial v_i^{(h-1)}} a_j^{(h-1)} = \delta_i^{(h-1)} a_j^{(h-2)}$$

$$\delta_i^{(h-2)} = \sum_k \delta_k^{(h)} w_{ki}^{(h)} \varphi^{(h-1)\prime}(v_i^{(h-1)})$$

При выводе формул предполагаем что $\frac{\partial a_k^{(l)}}{\partial w_{i,j}^{(l)}} = 0$ при $k \neq i$ (условие 1)

Движемся обратно через сеть и следуем цепочному правилу

$$\frac{\partial E}{\partial w_{ij}^{(h-2)}} = \sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial w_{ij}^{(h-2)}} = \sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial v_k^{(h)}} \frac{\partial v_k^{(h)}}{\partial w_{ij}^{(h-2)}}$$

$$\sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial v_k^{(h)}} \sum_l \frac{\partial v_k^{(h)}}{\partial a_l^{(h-1)}} \frac{\partial a_l^{(h-1)}}{\partial v_l^{(h-1)}} \frac{\partial v_l^{(h-1)}}{\partial a_i^{(h-2)}} \frac{\partial a_i^{(h-2)}}{\partial v_i^{(h-2)}} \frac{\partial v_i^{(h-2)}}{\partial w_{ij}^{(h-2)}}$$

$$\sum_k \delta_k^{(h)} \sum_l w_{kl}^{(h)} \varphi^{(h-1)'}(v_l^{(h-1)}) w_{li}^{(h-1)} \varphi^{(h-2)'}(v_i^{(h-2)}) a_j^{(h-3)}$$

$$\sum_l \sum_k \delta_k^{(h)} w_{kl}^{(h)} \varphi^{(h-1)'}(v_l^{(h-1)}) w_{li}^{(h-1)} \varphi^{(h-2)'}(v_i^{(h-2)}) a_j^{(h-3)}$$

$$\sum_l \sum_k \delta_k^{(h)} w_{kl}^{(h)} \varphi^{(h-1)'}(v_l^{(h-1)}) w_{li}^{(h-1)} \varphi^{(h-2)'}(v_i^{(h-2)}) a_j^{(h-3)}$$

$$\sum_l \delta_l^{(h-1)} w_{li}^{(h)} \varphi^{(h-2)'}(v_i^{(h-2)}) a_j^{(h-3)} = \delta_i^{(h-2)} a_j^{(h-3)}$$

Переобозначим l в k, получаем

$$\delta_i^{(h-2)} = \sum_k \delta_k^{(h-1)} w_{ki}^{(h-1)} \varphi^{(h-2)'}(v_i^{(h-2)})$$

из чего выводим общую формулу

$$\delta_i^{(s-1)} = \varphi^{(s-1)'}(v_i^{(s-1)}) \sum_k \delta_k^{(s)} w_{ki}^{(s)} \quad \text{и} \quad \delta_i^{(h)} = \frac{\partial E}{\partial v_i^{(h)}}$$

$(s \in \{1 \ldots h\})$

$$\frac{\partial E}{\partial w_{ij}^{(s)}} = \delta_i^{(s)} a_j^{(s-1)}$$

В матричном виде

$$\delta^{(h)} = \frac{\partial E}{\partial v^{(h)}} \qquad \delta^{(s)} = \varphi^{(s)'}(v^{(s)}) * \left(W^{(s+1)T} \delta^{(s+1)}\right)$$

поэлементное умножение

$$\frac{\partial E}{\partial W^{(s)}} = \delta^{(s)} a^{(s-1)T}$$

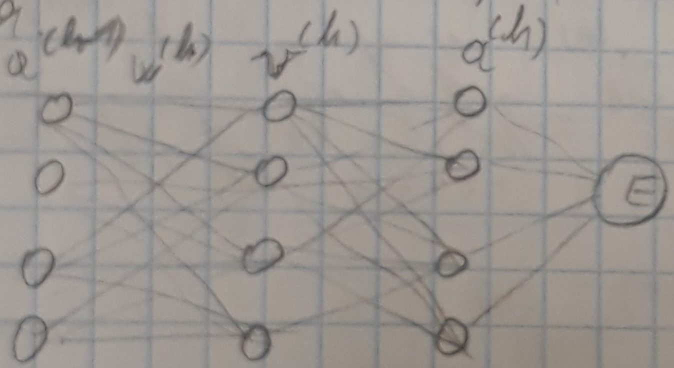
Если у нас больше одного входного примера

$$E_N(w) = \frac{1}{N}\sum_{n=1}^{N} E(w)$$

$$\frac{\partial E_N(w)}{\partial w^{(s)}} = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial E}{\partial w^{(s)}} = \sum_{n=1}^{N} \frac{1}{N}\delta^{(s,n)} a^{(s-1,n)^T}$$

$$\Delta w^{(s)} = \sum_{n=1}^{N} \frac{1}{N}\delta^{(s,n)} a^{(s-1,n)^T}$$

Если у нас softmax функция активации $\varphi^{(h)}(v_i^{(h)}) = \frac{e^{v_i^{(h)}}}{\sum_k e^{v_k^{(h)}}}$, видно что условие 1 не выполнено, так как её значение не зависит только от $v_i^{(h)}$, а от всех $v_k^{(h)}$ последнего слоя



Но $v_i^{(h)}$ зависит только от $w_{i\bullet}^{(h)}$, поэтому получаем одинаковую формулу:

$$\frac{\partial E}{\partial w_{ij}^{(h)}} = \sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial w_{ij}^{(h)}} = \sum_k \frac{\partial E}{\partial a_k^{(h)}} \sum_p \frac{\partial a_k^{(h)}}{\partial v_p^{(h)}} \frac{\partial v_p^{(h)}}{\partial w_{ij}^{(h)}}$$

$$= \sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial v_i^{(h)}} \frac{\partial v_i^{(h)}}{\partial w_{ij}^{(h)}} = \sum_k \frac{\partial E}{\partial a_k^{(h)}} \frac{\partial a_k^{(h)}}{\partial v_i^{(h)}} a_j^{(h-1)}$$

$$= \frac{\partial E}{\partial v_i^{(h)}} a_j^{(h-1)} = \delta_i a_j^{(h-1)}$$

Классификация k классов, при softmax функции активации и кросс-энтропийной функции ошибки, с использованием onehot кодирования:

Активации softmax представляют апостериорные вероятности что пример принадлежит данному классу

$$p(y^{(n)} \mid w, x^{(n)}) = \frac{e^{u_j}}{\sum_k e^{u_k}}$$

Хотим максимизировать функцию правдоподобия:

$$p(Y \mid w) = \prod_{n=1}^{N} \prod_{j=1}^{K} \left(\frac{e^{u_j}}{\sum_k e^{u_k}} \right)^{y_j^{(n)}}.$$

так получаем кросс-энтропийную функцию ошибки:

$$E_1(w) = -\frac{1}{N} \ln \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\frac{e^{u_j^{(h,n)}}}{\sum_k e^{u_k^{(h,n)}}} \right)^{y_j^{(n)}} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{K} y_j^{(n)} \frac{e^{u_j^{(h,n)}}}{\sum_k e^{u_k^{(h,n)}}}$$

$$= -\frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{K} \left(u_j^{(h,n)} - \ln \sum_k e^{u_k^{(h,n)}} \right) y_j^{(n)}$$

Ошибка при одном входном примере:

$$E(w) = \sum_{j=1}^{K} y_j \frac{e^{u_j^{(h)}}}{\sum_k e^{u_k^{(h)}}} = \sum_{j=1}^{K} y_j a_j^{(h)}$$

$$\frac{\partial E}{\partial u_i^{(h)}} = \sum_{\substack{j \neq i \\ j=1}}^{K} \left(y_j \frac{e^{u_j^{(h)}}}{\sum_k e^{u_k^{(h)}}} \right) + \left(\frac{y_i e^{u_i^{(h)}}}{\sum_k e^{u_k^{(h)}}} - y_i \right)$$

$$= \sum_{j=1}^{K} \frac{y_j e^{u_k^{(h)}}}{\sum_k e^{u_k^{(h)}}} - y_i = \frac{e^{u_i^{(h)}}}{\sum_k e^{u_k^{(h)}}} - y_i = o_i^{(h)} - y_i = \delta_i^{(h)}$$

Второй способ нахождения $\delta_i^{(n)}$.

$$\frac{\partial a_i}{\partial u_i} = \frac{e^{u_i}\sum_k e^{u_k} - (e^{u_i})^2}{(\sum_k e^{u_k})^2} = a_i(1-a_i)$$

$$\frac{\partial a_j}{\partial u_i} = -e^{u_j}\frac{e^{u_i}}{(\sum_k e^{u_k})^2} = -a_j a_i$$

$$\frac{\partial E}{\partial a_i} = \frac{\partial\left(-\sum_{j=1}^{k} y_j \ln a_j\right)}{\partial a_i} = -\sum_{j=1}^{k} y_j \frac{\partial \ln a_j}{\partial a_i} = -\frac{y_i}{a_i}$$

$$\frac{\partial E}{\partial u_i} = \sum_{j=1}^{k}\frac{\partial E}{\partial a_j}\frac{\partial a_j}{\partial u_i} = \sum_{j=1}^{k} -\frac{y_j}{a_j}\frac{\partial a_j}{\partial u_i} = \sum_{\substack{j\neq i\\ j=1}}^{k} y_j a_i - y_i(1-a_i)$$

Если $y_i = 1 \Rightarrow a_i - 1 = a_i - y_i$

Если $y_i = 0 \Rightarrow a_i = a_i - 0 = a_i - y_i$

$$\frac{\partial E}{\partial u_i} = a_i - y_i$$

$$\delta^{(n)} = a^{(n)} - y.$$