***Обратное распространение ошибки.***

 В общем случае:



Далее везде $z\_{i}^{(l)}$ – выход *i*-го нейрона в слое сномером *l*, $x\_{i}^{l}=f\left(z\_{i}^{l}\right)$ – результат применения активационной функции к выходу соответствующего нейрона или то, что поступает на вход след. слоя.

Целевая функция:

$$F\left(w, b\right)=\sum\_{i}^{}δ\left(t\_{i}-f\left(z\_{i}^{(3)}\right)\right)$$

Выход сети:

$$f\left(z\_{i}^{(3)}\right)=f\left(\sum\_{j}^{}w\_{ij}^{(3)}x\_{j}^{\left(2\right)}+b\_{i}^{(3)}\right)$$

$$z\_{i}^{(l)}=\sum\_{j}^{}w\_{ij}^{(l)}x\_{j}^{\left(l-1\right)}+b\_{i}^{(l)}$$

Здесь $w\_{ij}^{(l)} и b\_{i}^{(l)}$ веса и смещения текущего слоя, а $x\_{i}^{\left(l-1\right)}$ выходные данные предыдущего слоя, индекс $i$ обозначает номер нейрона в текущем слое.

$$x\_{j}^{\left(n\right)}=σ\left(\sum\_{m}^{}w\_{jm}^{(h)}x\_{m}^{(n)}+b\_{j}^{(h)}\right)$$

 Таким образом, соединяя входы и выходы скрытых слоёв, мы можем наращивать сложность нейросети.

 В общем случае результат для многоуровневой нейросети можно представить в следующей форме :

$$y=f\left(W^{(L)}f\left(W^{(L-1)}f\left(…\right)+b^{(L-1)}\right)+b^{(L)}\right)$$

***В этой и следующих формулах верхний индекс обозначает уровень нейросети, выходной уровень - последний.***

Для обучения сети нам нужна целевая функция или функция ошибок:

$$F\left(w, b\right)=U\left(x\_{c}\right)$$

где $x\_{c}=f\left(z\_{c}^{(L)}\right)$ – выходные данные нейросети, *с* – номер нейрона, $f$ – активационная функция.

Часто используемые формулы:

$$z\_{i}^{(l)}=\sum\_{j}^{}w\_{ij}^{(l)}f\left(z\_{j}^{(l-1)}\right)+b\_{i}^{(l)}$$

$$\frac{∂z\_{i}^{(l)}}{∂b\_{k}^{(l-n)}}=\left\{\begin{array}{c}δ\_{ik} если n=0\\\sum\_{j}^{}w\_{ij}^{(l)}f'\left(z\_{j}^{(l-1)}\right)\frac{∂z\_{j}^{(l-1)}}{∂b\_{k}^{(l-n)}}\end{array}\right.$$

Производная по параметрам (весам и смещениям) для последнего слоя *L* имеет вид:

$$\frac{∂F\left(w,b\right)}{∂b\_{k}^{(L)}}=\sum\_{c}^{}\frac{∂U\left(x\_{c}\right)}{∂x\_{c}}\frac{∂f\left(z\_{c}^{(L)}\right)}{∂b\_{k}^{(L)}}$$

L – последний слой, индекс c – перебор выходных нейронов.

$$=\sum\_{c}^{}\frac{∂U\left(x\_{c}\right)}{∂x\_{c}}f'\left(z\_{c}^{(L)}\right)\frac{∂z\_{c}^{(L)}}{∂b\_{k}^{(L)}}=\frac{∂U\left(x\_{k}\right)}{∂x\_{k}}f\_{z}'\left(z\_{k}^{(L)}\right)$$

напомним, что $x\_{k}=f\left(z\_{k}^{(L)}\right)$

$$\frac{∂F\left(w,b\right)}{∂b\_{k}^{(L)}}=ε\_{k}^{(L)}$$

$$ε\_{k}^{(L)}=\frac{∂U\left(x\_{k}\right)}{∂x\_{k}}f\_{z}'\left(z\_{k}^{(L)}\right)$$

Для предпоследнего слоя L-1:

$$\frac{∂F\left(w,b\right)}{∂b\_{k}^{(L-1)}}=\sum\_{c}^{}\frac{∂U\left(x\_{c}\right)}{∂x\_{c}}\frac{∂f\left(z\_{c}^{(L)}\right)}{∂b\_{k}^{(L-1)}}=\sum\_{c}^{}\frac{∂U\left(x\_{c}\right)}{∂x\_{c}}\frac{∂f\left(z\_{c}^{(L)}\right)}{∂z\_{k}^{(L)}}\frac{\frac{∂f\left(z\_{c}^{(L)}\right)}{∂b\_{k}^{(L-1)}}}{∂b\_{k}^{(L-1)}}=\sum\_{c}^{}ε\_{c}^{(L)}\frac{∂z\_{c}^{(L)}}{∂b\_{k}^{(L-1)}}=\sum\_{c}^{}ε\_{c}^{(L)}\sum\_{i}^{}w\_{ci}^{(L)}f'\left(z\_{i}^{(L-1)}\right)\frac{∂z\_{i}^{(L-1)}}{∂b\_{k}^{(L-1)}}=\sum\_{c}^{}ε\_{c}^{(L)}w\_{ck}^{(L)}f'\left(u\_{k}^{(L-1)}\right)$$

$$\frac{∂F\left(w,b\right)}{∂b\_{k}^{(L-1)}}=ε\_{k}^{(L-1)}$$

$$ε\_{k}^{(L-1)}=\sum\_{c}^{}ε\_{c}^{(L)}w\_{ck}^{(L)}f'\left(z\_{k}^{(L-1)}\right)$$

Для слоя L-2:

$$\frac{∂F\left(w,b\right)}{∂b\_{k}^{(L-2)}}=\sum\_{i}^{}ε\_{i}^{(L-1)}\frac{∂z\_{i}^{(L-1)}}{∂b\_{k}^{(L-2)}}=\sum\_{i}^{}ε\_{i}^{(L-1)}\sum\_{j}^{}w\_{ij}^{(L-1)}f'\left(z\_{j}^{(L-2)}\right)\frac{∂z\_{j}^{(L-2)}}{∂b\_{k}^{(L-2)}}=\sum\_{i}^{}ε\_{i}^{(L-1)}w\_{ik}^{(L-1)}f'\left(z\_{k}^{(L-2)}\right)$$

$$\frac{∂F\left(w,b\right)}{∂b\_{k}^{(L-2)}}=ε\_{k}^{(L-2)}$$

$$ε\_{k}^{(L-2)}=\sum\_{i}^{}ε\_{i}^{(L-1)}w\_{ik}^{(L-1)}f'\left(z\_{k}^{(L-2)}\right)$$

Попробуем записать производную по весовым коэффициентам для произвольного слоя:

$$\frac{∂F\left(w,b\right)}{∂w\_{kl}^{(L-n)}}=\sum\_{i}^{}ε\_{i}^{(L-n+1)}\sum\_{j}^{}w\_{ij}^{(L-n+1)}f'\left(z\_{j}^{(L-n)}\right)\frac{∂z\_{j}^{(L-n)}}{∂w\_{kl}^{(L-n)}}=\sum\_{i}^{}ε\_{i}^{(L-n+1)}\sum\_{j}^{}w\_{ij}^{(L-n+1)}f'\left(z\_{j}^{(L-n)}\right)\frac{∂}{∂w\_{kl}^{(L-n)}}\left(\sum\_{s}^{}w\_{js}^{(L-n)}f\left(z\_{s}^{(L-n-1)}\right)+b\_{i}^{(L-n)}\right)=\sum\_{i}^{}ε\_{i}^{(L-n+1)}w\_{ik}^{(L-n+1)}f'\left(z\_{k}^{(L-n)}\right)f\left(z\_{l}^{(L-n-1)}\right)$$

В итоге получаем:

$$\frac{∂F\left(w,b\right)}{∂w\_{kl}^{\left(L-n\right)}}=\sum\_{i}^{}ε\_{i}^{\left(L-n+1\right)}w\_{ik}^{\left(L-n+1\right)}f^{'\left(u\_{k}^{\left(L-n\right)}\right)}∙\left\{\begin{array}{c}x\_{l} -входные данные\\f\left(u\_{l}^{\left(L-n-1\right)}\right) -скрытый слой\end{array}\right. $$

Выходной слой:

$$\frac{∂F\left(w,b\right)}{∂b\_{k}^{(L)}}=ε\_{k}^{(L)}$$

$$ε\_{k}^{(L)}=\frac{∂U\left(x\_{k}\right)}{∂x\_{k}}f'\left(z\_{k}^{(L)}\right)$$

Скрытый слой:

$$\frac{∂F\left(w,b\right)}{∂b\_{k}^{\left(L-n\right)}}=ε\_{k}^{\left(L-n\right)}$$

$$ε\_{k}^{(L-n)}=\sum\_{i}^{}ε\_{i}^{(L-n+1)}w\_{ik}^{(L-n+1)}f'\left(z\_{k}^{(L-n)}\right)$$

 Приведённый способ вычисления производных целевой функции по параметрам различных слоёв нейросети называется метод обратного распространения ошибки, так как в качестве целевой функции обычно применяется функция ошибок.