***Обратное распространение ошибки.***



Обозначим

$l$ – номер слоя

$w\_{l}, b\_{l}$ – параметры слоя

$x^{(l)}$ – входные данные слоя

$z^{(l+1)}, x^{(l+1)}=σ\left(z^{(l+1)}\right)$ – выходные данные слоя

Пусть L – число слоёв нейросети. Слой с номером L – выходной слой нейросети.

Целевая функция:

$$F\left(w, b\right)=\frac{1}{2}\left(t-x^{(L+1)}\right)^{2}$$

 Результат обработки данных в скрытом слое:

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

$$x\_{i}^{\left(l\right)}=σ\left(z\_{i}^{l}\right)$$

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

$σ(…)$ – активационная функция. Ранее в качестве активационно мы использовали индикаторную функцию $I\left(x\right)$. Здесь и далее используем экспоненциальную сигмоиду:

$$σ\left(x\right)=\frac{1}{1+e^{-ax}}$$

$$\frac{dσ\left(x\right)}{dx}=aσ\left(x\right)(1-σ\left(x\right))$$

Заметим, что $σ(x)→I\left(x\right)$.

Выход сети можно представить через результат работы предыдущего слоя

$$x\_{0}^{(L+1)}=σ\left(\sum\_{j}^{}w\_{0j}^{(L)}x\_{j}^{\left(L\right)}+b\_{0}^{(L)}\right)$$

 Активационная функция выходного слоя зависит от типа решаемой задачи. Для задачи классификации данных на два класса используем сигмоиду.

Для удобства изложения результат работы сети (как илюбого скрытого слоя) будем представлять в виде вектора. В рассматриваемом случае выходной вектор имеет одну компоненту $x\_{0}^{(L)}$.

 Соединяя входы и выходы слоёв мы можем наращивать сложность нейросети. В рассматриваемом случае результат работы нейросети можно представить в следующей форме :

$$x^{(L+1)}=σ\left(W^{(L)}σ\left(W^{(L-1)}σ\left(…\right)+B^{(L-1)}\right)+B^{(L)}\right)$$

здесь *W* и *B* обозначают матрицу весов и вектор смещений соответствующего слоя сети.

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

Обучение сети сводится к минимизации целевой функции или функции ошибок:

$$F\left(w, b\right)=\frac{1}{2}\left(x^{\left(L+1\right)}-t\right)^{2}$$

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=\sum\_{i}^{}\frac{∂F}{∂x\_{i}^{(L+1)}}\frac{∂x\_{i}^{(L+1)}}{∂b\_{k}^{\left(L-l\right)}}=\sum\_{i}^{}\frac{∂F}{∂x\_{i}^{(L+1)}}\frac{∂σ\left(z\_{i}^{(L+1)}\right)}{∂b\_{k}^{\left(L-l\right)}}$$

$$=\sum\_{i}^{}\frac{∂F}{∂x\_{i}^{(L+1)}}\frac{∂σ\left(z\_{i}^{(L+1)}\right)}{∂z\_{i}^{\left(L+1\right)}}\frac{∂}{∂b\_{k}^{\left(L-l\right)}}z\_{i}^{\left(L+1\right)}$$

$$=\sum\_{i}^{}\frac{∂F}{∂x\_{i}^{(L+1)}}aσ\left(z\_{i}^{(L+1)}\right)\left(1-σ\left(z\_{i}^{(L+1)}\right)\right)\frac{∂}{∂b\_{k}^{\left(L-l\right)}}\left(\sum\_{j}^{}w\_{ij}^{(L)}x\_{j}^{\left(L\right)}+b\_{i}^{(L)}\right)$$

$$=\sum\_{i}^{}\frac{∂F}{∂x\_{i}^{(L+1)}}ax\_{i}^{(L+1)}\left(1-x\_{i}^{(L+1)}\right)\frac{∂}{∂b\_{k}^{\left(L-l\right)}}\left(\sum\_{j}^{}w\_{ij}^{(L)}x\_{j}^{\left(L\right)}+b\_{i}^{(L)}\right)$$

если *l*=0

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L\right)}}=\sum\_{i}^{}\frac{∂F}{∂x\_{i}^{(L+1)}}ax\_{i}^{(L+1)}\left(1-x\_{i}^{(L+1)}\right)δ\_{ik}$$

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

при *l>*0

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=\sum\_{i}^{}\frac{∂F}{∂x\_{i}^{(L+1)}}ax\_{i}^{(L+1)}\left(1-x\_{i}^{(L+1)}\right)\sum\_{j}^{}w\_{ij}^{(L)}\frac{∂x\_{j}^{\left(L\right)}}{∂b\_{k}^{\left(L-l\right)}}$$

$$=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}w\_{ij}^{(L)}\frac{∂x\_{j}^{\left(L\right)}}{∂b\_{k}^{\left(L-l\right)}}$$

$$=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}w\_{ij}^{(L)}\frac{∂σ\left(z\_{j}^{\left(L\right)}\right)}{∂b\_{k}^{\left(L-l\right)}}=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}w\_{ij}^{(L)}\frac{∂σ\left(z\_{j}^{\left(L\right)}\right)}{∂z\_{j}^{\left(L\right)}}\frac{∂z\_{j}^{\left(L\right)}}{∂b\_{k}^{\left(L-l\right)}}$$

$$=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}w\_{ij}^{(L)}\frac{∂σ\left(z\_{j}^{\left(L\right)}\right)}{∂z\_{j}^{\left(L\right)}}\frac{∂}{∂b\_{k}^{\left(L-l\right)}}\left(\sum\_{j\_{1}}^{}w\_{jj\_{1}}^{(L-1)}x\_{j\_{1}}^{(L-1)}+b\_{j}^{(L-1)}\right)$$

обозначим

$$U\_{ij}^{(l)}=w\_{ij}^{(l)}\frac{∂σ\left(z\_{j}^{\left(l\right)}\right)}{∂z\_{j}^{\left(l\right)}}$$

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}U\_{ij}^{(L)}\frac{∂}{∂b\_{k}^{\left(L-l\right)}}\left(\sum\_{j\_{1}}^{}w\_{jj\_{1}}^{(L-1)}x\_{j\_{1}}^{(L-1)}+b\_{j}^{(L-1)}\right)$$

если *l*=1

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-1\right)}}=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}U\_{ij}^{(L)}δ\_{kj}=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}U\_{ik}^{(L)}=\left(\frac{∂F\left(w, b\right)}{∂b^{(L)}}⋅U^{(L)}\right)\_{k}$$

при *l>*1

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}U\_{ij}^{(L)}\sum\_{j\_{1}}^{}w\_{jj\_{1}}^{(L-1)}\frac{∂x\_{j\_{1}}^{(L-1)}}{∂b\_{k}^{\left(L-l\right)}}$$

$$=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}U\_{ij}^{(L)}\sum\_{j\_{1}}^{}w\_{jj\_{1}}^{(L-1)}\frac{∂σ\left(z\_{j\_{1}}^{(L-1)}\right)}{∂b\_{k}^{\left(L-l\right)}}$$

$$=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}U\_{ij}^{(L)}\sum\_{j\_{1}}^{}w\_{jj\_{1}}^{(L-1)}\frac{∂σ\left(z\_{j\_{1}}^{(L-1)}\right)}{∂z\_{j\_{1}}^{(L-1)}}\frac{∂z\_{j\_{1}}^{(L-1)}}{∂b\_{k}^{\left(L-l\right)}}$$

$$=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}U\_{ij}^{(L)}\sum\_{j\_{1}}^{}w\_{jj\_{1}}^{(L-1)}\frac{∂σ\left(z\_{j\_{1}}^{(L-1)}\right)}{∂z\_{j\_{1}}^{(L-1)}}\frac{∂}{∂b\_{k}^{\left(L-l\right)}}\left(\sum\_{j\_{2}}^{}w\_{j\_{1}j\_{2}}^{(L-2)}x\_{j\_{2}}^{(L-2)}+b\_{j\_{1}}^{(L-2)}\right)$$

если *l*=2

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}U\_{ij}^{(L)}w\_{jk}^{(L-1)}\frac{∂σ\left(z\_{k}^{(L-1)}\right)}{∂z\_{k}^{(L-1)}}$$

$$=\sum\_{i}^{}\frac{∂F\left(w, b\right)}{∂b\_{i}^{\left(L\right)}}\sum\_{j}^{}U\_{ij}^{(L)}U\_{jk}^{(L-1)}$$

в матричной форме:

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

для произвольного значения *l*

 $$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=\left(\frac{∂F\left(w, b\right)}{∂b^{(L)}}⋅U^{(L)}⋅U^{(L-1)}⋅…⋅U^{(L-l+1)}\right)\_{k}$$

 Аналогичные рассуждения для производной по весу $w\_{km}^{(L-l)}$. Отличие при нахождении производной

$$\frac{∂z\_{j\_{1}}^{(L-l)}}{∂w\_{km}^{\left(L-l\right)}}=\frac{∂}{∂w\_{km}^{\left(L-l\right)}}\left(\sum\_{j\_{2}}^{}w\_{j\_{1}j\_{2}}^{(L-l)}x\_{j\_{2}}^{(L-l)}+b\_{j\_{1}}^{(L-l)}\right)$$

$$=\sum\_{j\_{2}}^{}δ\_{kj\_{1}}δ\_{mj\_{2}}x\_{j\_{2}}^{(L-l)}=δ\_{kj\_{1}}x\_{m}^{(L-l)}$$

Таким образом:

$$\frac{∂F\left(w, b\right)}{∂w\_{km}^{\left(L-l\right)}}=\left(\frac{∂F\left(w, b\right)}{∂b^{(L)}}⋅U^{(L)}⋅U^{(L-1)}⋅…⋅U^{(L-l+1)}\right)\_{k}x\_{m}^{(L-l)}$$

С учётом явного вида функции ошибок (потерь)

$$\frac{∂F}{∂x\_{i}^{(L+1)}}=\frac{1}{2}\frac{∂}{∂x\_{i}^{(L+1)}}\left(x^{\left(L+1\right)}-t\right)^{2}=\left(x^{(L+1)}-t\right)\_{i}$$

окончательно получаем:

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=\left(\frac{∂F\left(w, b\right)}{∂b^{(L)}}⋅U^{(L)}⋅U^{(L-1)}⋅…⋅U^{(L-l+1)}\right)\_{k}$$

или

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

Для производной по весу

$$\frac{∂F\left(w, b\right)}{∂w\_{km}^{\left(L-l\right)}}=\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}x\_{m}^{(L-l)}$$

где

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

$$U\_{ij}^{(l)}=w\_{ij}^{(l)}\frac{∂σ\left(z\_{j}^{\left(l\right)}\right)}{∂z\_{j}^{\left(l\right)}}=w\_{ij}^{(l)}aσ\left(z\_{j}^{\left(l\right)}\right)\left(1-σ\left(z\_{j}^{\left(l\right)}\right)\right)=w\_{ij}^{(l)}ax\_{j}^{\left(l\right)}\left(1-x\_{j}^{\left(l\right)}\right)$$