***Несколько выходов сети (нейронов в выходном слое больше одного)***

 Ранее мы использовали нейросеть для классифиции данных на 2 класса. Использовали нейросеть с одним нейроном в выходном слое:



Значение 0 или 1 на выходе сети указывало принадлежность объекта первому или второму классу.

 Используем сеть с несколькими выходами (в данном случае с двумя):



В этом случае номер выхода со значением 1 указывает номер класса.

 Для обучения сети можно использовать целевую функцию вида:

$$F\left(w, b\right)=\sum\_{train}^{}\sum\_{n}^{}I\left(t\_{n}-x\_{n}^{(L)}=0\right)$$

здесь $L$ – номер последнего слоя, n – номер нейрона в выходном слое

Значение функция $F\left(w, b\right)$ равно числу правильно классифицированных точек в тренирочном наборе данных.

 Правильные значения весовых множетелей *w* и смещения *b* обеспечивают максимальному значению целевой функции $F\left(w, b\right)$.

 Как и прежде, заменим индикаторную функцию на гладкую, дифференцируемую функцию:

$$I\left(z>0\right)\rightarrow σ\left(z\right)=\frac{1}{1+e^{-az}} \frac{dσ(z)}{dz}=aσ\left(z\right)\left(1-σ\left(z\right)\right)$$

$$I\left(z=0\right)\rightarrow δ\left(z\right)=e^{-az^{2}} \frac{dδ(z)}{dz}=-2azδ(z)$$



В этом случае целевая функция имеет вид:

$$F\left(w, b\right)=\sum\_{train}^{}\sum\_{n}^{}δ\left(t\_{n}-x\_{n}^{(L)}\right)=\sum\_{train}^{}\sum\_{n}^{}exp\left(-a\left(t\_{n}-x\_{n}^{(L)}\right)^{2}\right)$$

$$\frac{1}{\left(t\_{n}-x\_{n}^{(L)}\right)^{2}+d^{2}}$$

***Softmax – активационная функция***

 Выход сети с максимальным значением указывает номер класса, которому принадлежат входные данные. Для того чтобы увеличить значение на «правильном» выходе и придать выходным значениям смысл вероятности вхождения объекта в тот или иной класс в качестве активационной функции выходного слоя используется Softmax функция. Например:

$$x\_{n}^{(L)}=S\left(z\_{n}^{(L)}\right)={exp\left(α∙z\_{n}^{(L)}\right)}/{\sum\_{i}^{}exp\left(α∙z\_{i}^{(L)}\right)}$$

$$\sum\_{n}^{}x\_{n}^{(L)}=1$$

$L$ – номер последнего слоя,

n – номер нейрона в выходном слое.

 Возвращаемся к обратному распространению ошибки.

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=\frac{∂}{∂b\_{k}^{\left(L-l\right)}}\sum\_{n}^{}δ\left(t\_{n}-x\_{n}^{(L)}\right)=\frac{∂}{∂b\_{k}^{\left(L-l\right)}}\sum\_{n}^{}exp\left(-a\left(t\_{n}-x\_{n}^{(L)}\right)^{2}\right)$$

$$=-2a\sum\_{n}^{}\left(t\_{n}-x\_{n}^{(L)}\right)exp\left(-a\left(t\_{n}-x\_{n}^{(L)}\right)^{2}\right)\frac{∂}{∂b\_{k}^{\left(L-l\right)}}\left(t\_{n}-x\_{n}^{(L)}\right)$$

$$=2a\sum\_{n}^{}\left(t\_{n}-x\_{n}^{\left(L\right)}\right)δ\left(t\_{n}-x\_{n}^{\left(L\right)}\right)\frac{∂x\_{n}^{\left(L\right)}}{∂b\_{k}^{\left(L-l\right)}}$$

$$=2a\sum\_{n}^{}\left(t\_{n}-x\_{n}^{(L)}\right)δ\left(t\_{n}-x\_{n}^{(L)}\right)\frac{∂S\left(z\_{n}^{(L)}\right)}{∂b\_{k}^{\left(L-l\right)}}$$

$$=2a\sum\_{n}^{}\left(t\_{n}-x\_{n}^{(L)}\right)δ\left(t\_{n}-x\_{n}^{(L)}\right)\frac{∂S\left(z\_{n}^{(L)}\right)}{∂z\_{n}^{(L)}}\frac{∂z\_{n}^{(L)}}{∂b\_{k}^{\left(L-l\right)}}$$

-------------------------------------------------------------------------

$$\frac{∂S\left(z\_{n}\right)}{∂z\_{n}}=\frac{∂}{∂z\_{n}}\left({exp\left(α∙z\_{n}\right)}/{\sum\_{i}^{}exp\left(α∙z\_{i}\right)}\right)$$

$$=α{exp\left(α∙z\_{n}\right)}/{\sum\_{i}^{}exp\left(α∙z\_{i}\right)}-{exp\left(α∙z\_{n}\right)}/{\left(\sum\_{i}^{}exp\left(α∙z\_{i}\right)\right)^{2}}\frac{∂}{∂z\_{n}}\left(\sum\_{i}^{}exp\left(α∙z\_{i}\right)\right)$$

$$=αS\left(z\_{n}\right)-{S\left(z\_{n}\right)}/{\sum\_{i}^{}exp\left(α∙z\_{i}\right)}\left(α\sum\_{i}^{}exp\left(α∙z\_{i}\right)δ\_{ni}\right)$$

$$=αS\left(z\_{n}\right)\left(1-S\left(z\_{n}\right)\right)=αx\_{n}\left(1-x\_{n}\right)$$

-------------------------------------------------------------------------

$$\frac{∂F\left(w, b\right)}{∂b\_{k}^{\left(L-l\right)}}=2aα\sum\_{n}^{}\left(t\_{n}-x\_{n}^{(L)}\right)δ\left(t\_{n}-x\_{n}^{(L)}\right)S\left(z\_{n}^{(L)}\right)\left(1-S\left(z\_{n}^{(L)}\right)\right)\frac{∂z\_{n}^{(L)}}{∂b\_{k}^{\left(L-l\right)}}$$

$$=2aα\sum\_{n}^{}\left(t\_{n}-x\_{n}^{(L)}\right)δ\left(t\_{n}-x\_{n}^{(L)}\right)x\_{n}^{(L)}\left(1-x\_{n}^{(L)}\right)\frac{∂z\_{n}^{(L)}}{∂b\_{k}^{\left(L-l\right)}}$$

 Далее всё как при рассмотрении обратного распространения ошибки, с добавлением множителя

$$\left(t\_{n}-x\_{n}^{(L)}\right)⟹\left(t\_{n}-x\_{n}^{(L)}\right)δ\left(t\_{n}-x\_{n}^{(L)}\right)$$

и заменой активационной функции

$$σ\left(z\_{n}^{(L)}\right)⟹S\left(z\_{n}^{(L)}\right)$$

для выходного слоя.