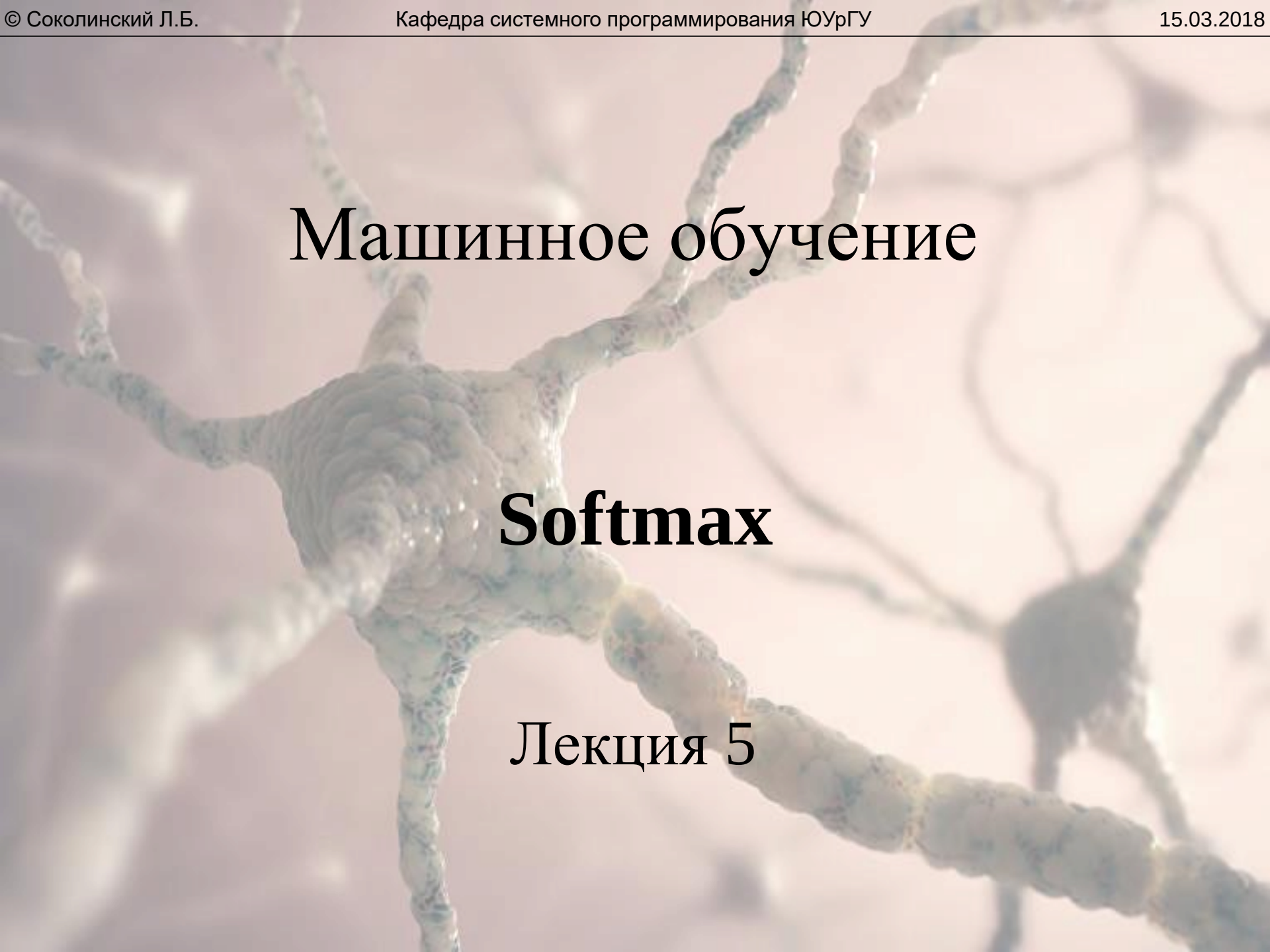


Машинное обучение

Softmax

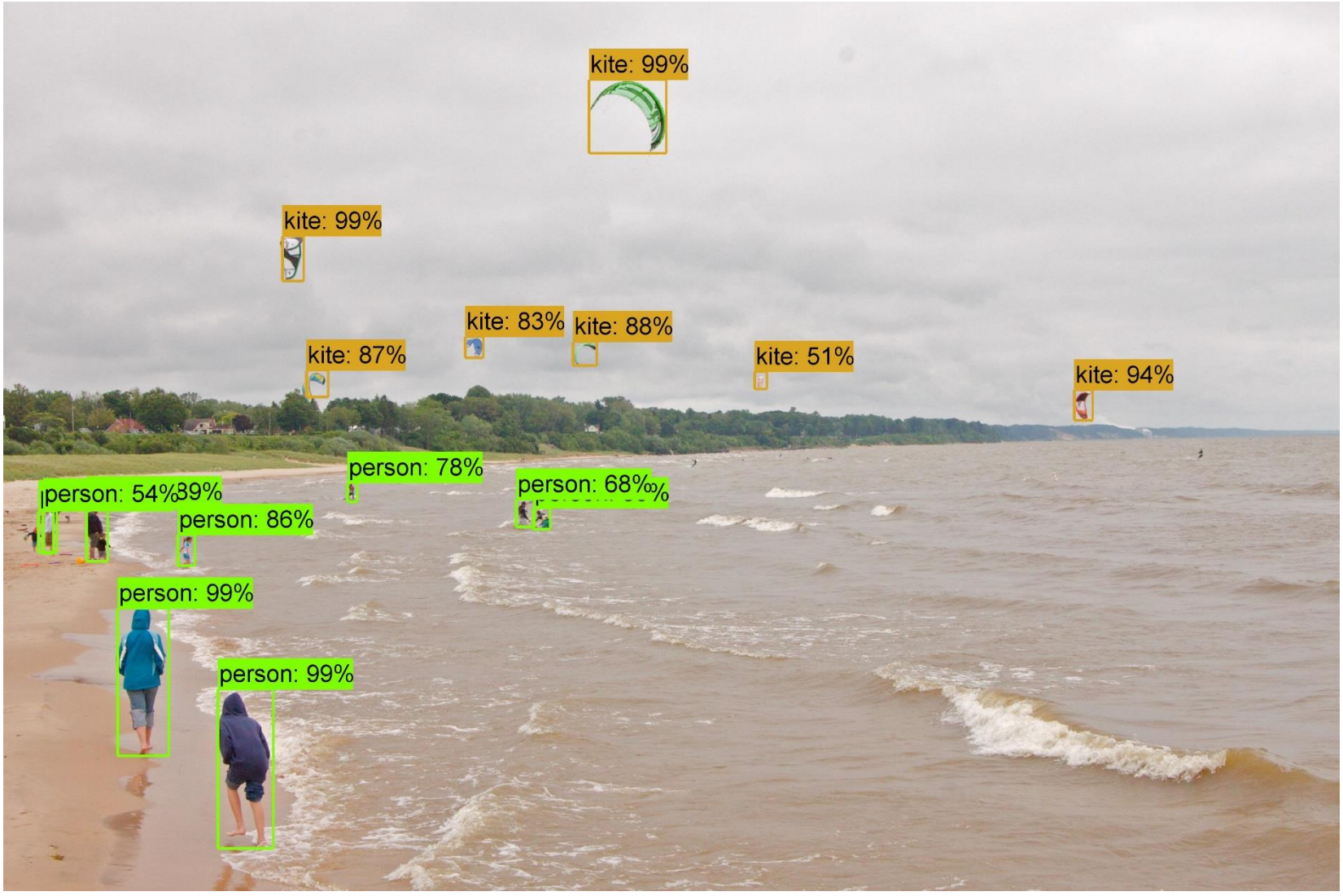
Лекция 5

На выходе нейронной сети выдается распределение вероятностей

$$0 \leq y_j \leq 1$$

$$\sum_{j=1}^{n} y_j = 1$$

Пример работы нейронной сети с распределением вероятностей



Использование функции *softmax* в качестве функции активации выходного слоя

n – количество нейронов в выходном слое L

m – количество нейронов в слое L-1

$$z_j^L = \sum_{k=1}^{m} w_{jk}^L a_k^{L-1} + b_j^L$$

Для выходного слоя L применяется особая функция активации *softmax*:

$$a_j^L = \frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}$$

Свойства функции *softmax*

$$1)\ 0 < \frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}} < 1 \Rightarrow 0 < a_j^L < 1$$

$$2)\ \sum_{j=1}^{n} a_j^L = \sum_{j=1}^{n} \frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}} = \frac{\sum_{j=1}^{n} e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}} = 1$$

n – количество нейронов в выходном слое L

m – количество нейронов в слое L-1

Выходной слой softmax задает распределение вероятностей

- Сумма всех выходных сигналов равна 1
- Выходной сигнал a_j^L j-того нейрона интерпретируется как вероятность того, что правильный ответ есть j
- Например, в задаче распознавания рукописных цифр значение a_4^L интерпретируется как вероятность того, что на вход сети подана цифра 4

Использование функции перекрестной энтропии в качестве функции потерь

$$C = -\sum_j y_j \ln a_j^L$$

Свойства функции потерь:

1) $C \geq 0$

2) Минимум C достигается при $\vec{a}^L = \vec{y}$:

$$\min_{\vec{a}^L} C = -\sum_j y_j \ln y_j$$

Доказательство $C \geq 0$

$$\left.\begin{array}{l} y_j \geq 0 \\ 0 < a_j^L < 1 \Rightarrow \ln a_j^L < 0 \end{array}\right\} \Rightarrow -y_j \ln a_j^L \geq 0 \Rightarrow$$

$$C = -\sum_j y_j \ln a_j^L \geq 0$$

Минимум достигается при $\vec{a}^L = \vec{y}$

y_1	0.5
y_2=(1-y_1)*2/3	0.33
y_3=(1-y_1)/3	0.17

min= 1.01

a_1	a_2=(1-a_1)*2/3	a_3=(1-a_1)/3	C=-(y_1*LN(a_1)+y_2*LN(a_2)+y_3*LN(a_3))
0.1	0.60	0.30	1.52
0.2	0.53	0.27	1.23
0.3	0.47	0.23	1.10
0.4	0.40	0.20	1.03
0.5	**0.33**	**0.17**	**1.01**
0.6	0.27	0.13	1.03
0.7	0.20	0.10	1.10
0.8	0.13	0.07	1.23
0.9	0.07	0.03	1.52

Вычисление ошибки δ_i^L для выходного слоя

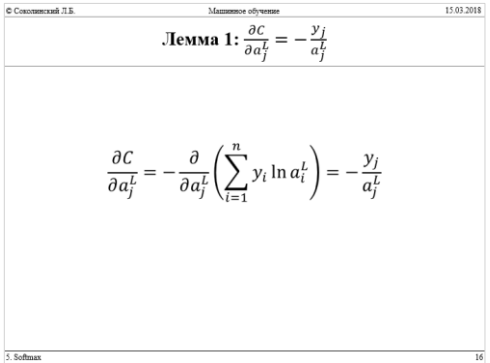
$$\delta_i^L = \frac{\partial C}{\partial z_i^L} = \sum_{j=1}^{n} \frac{\partial C}{\partial a_j^L} \cdot \frac{\partial a_j^L}{\partial z_i^L} = \sum_{j=1}^{n} -\frac{y_j}{a_j^L} \cdot \frac{\partial a_j^L}{\partial z_i^L} = -\sum_{j=1}^{n} \frac{y_j}{a_j^L} \cdot \frac{\partial a_j^L}{\partial z_i^L}$$

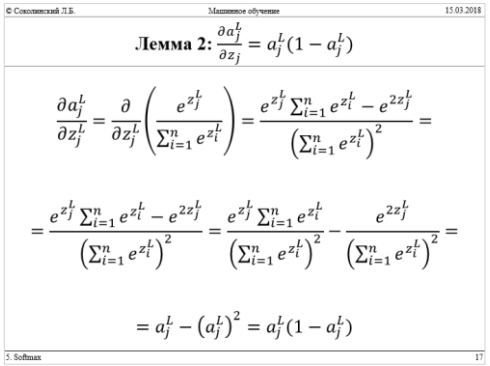
$$= -\frac{y_i}{a_i^L} \cdot \frac{\partial a_i^L}{\partial z_i^L} - \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{y_j}{a_j^L} \cdot \frac{\partial a_j^L}{\partial z_i^L}$$

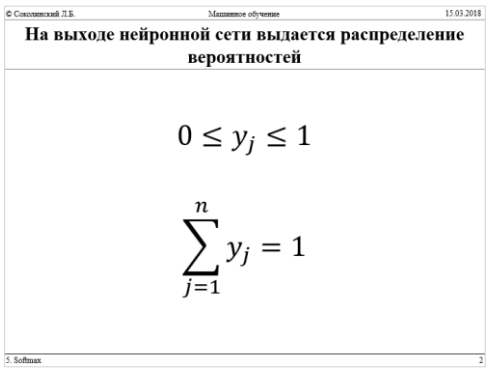
$$= -\frac{y_i}{a_i^L} a_i^L \left(1 - a_i^L\right) - \sum_{\substack{j=1 \\ j \neq i}}^{n} -\frac{y_j}{a_j^L} \cdot a_i^L a_j^L = -y_i + y_i a_i^L + \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j a_i^L$$

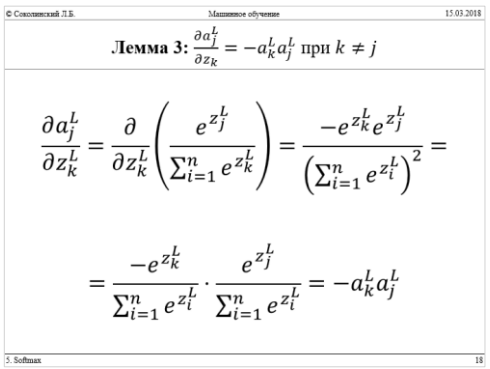
$$= -y_i + \sum_{j=1}^{n} y_j a_i^L = -y_i + a_i^L \sum_{j=1}^{n} y_j$$

$$= -y_i + a_i^L = a_i^L - y_i$$









Проблема замедления отсутствует, так как:

$$\frac{\partial C}{\partial b_i^L} = a_i^L - y_i$$

и

$$\frac{\partial C}{\partial w_{ik}^L} = a_k^{L-1}\left(a_i^L - y_i\right)$$

Доказательство $\frac{\partial C}{\partial b_i^L} = a_i^L - y_i$

$$\frac{\partial C}{\partial b_i^L} = \frac{\partial C}{\partial z_i^L} \cdot \frac{\partial z_i^L}{\partial b_i^L} = \frac{\partial C}{\partial z_i^L} \cdot 1 = \frac{\partial C}{\partial z_i^L} = a_i^L - y_i$$

Использование функции перекрестной энтропии в качестве функции потерь

$$C = -\sum_j y_j \ln a_j^L$$

Свойства функции потерь:

1) $C \geq 0$

2) Минимум C достигается при $\vec{a}^L = \vec{y}$:

$$\min_{\vec{a}^L} C = -\sum_j y_j \ln y_j$$

Использование функции *softmax* в качестве функции активации выходного слоя

n – количество нейронов в выходном слое L

m – количество нейронов в слое L-1

$$z_j^L = \sum_{k=1}^{m} w_{jk}^L a_k^{L-1} + b_j^L$$

Для выходного слоя L применяется особая функция активации *softmax*:

$$a_j^L = \frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}$$

Лемма 4: $\frac{\partial C}{\partial z_i^L} = a_i^L - y_i$

$$\frac{\partial C}{\partial z_i^L} = -\frac{\partial}{\partial z_i^L}\left(\sum_{j=1}^{n} y_j \ln a_j^L\right) = -\sum_{j=1}^{n} y_j \frac{\partial(\ln a_j^L)}{\partial z_i^L} = -\sum_{j=1}^{n} y_j \frac{\partial\left(\ln\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)\right)}{\partial z_i^L} = -\sum_{j=1}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L}$$

$$= -y_i \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_i^L}} \cdot \frac{\partial\left(\frac{e^{z_i^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L} - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L}$$

$$= -y_i \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_i^L}} \cdot \frac{e^{z_i^L}\sum_{i=1}^{n} e^{z_i^L} - e^{z_i^L}e^{z_i^L}}{\left(\sum_{i=1}^{n} e^{z_i^L}\right)^2} - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L} = -y_i \frac{\sum_{i=1}^{n} e^{z_i^L} - e^{z_i^L}}{\sum_{i=1}^{n} e^{z_i^L}} - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L}$$

$$= -y_i(1 - a_i^L) - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L} = -y_i(1 - a_i^L) - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{-e^{z_j^L}e^{z_i^L}}{\left(\sum_{i=1}^{n} e^{z_i^L}\right)^2} = -y_i(1 - a_i^L) - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{-e^{z_i^L}}{\sum_{i=1}^{n} e^{z_i^L}}$$

$$= -y_i(1 - a_i^L) - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j(-a_i^L) = -y_i(1 - a_i^L) + a_i^L \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j = -y_i + y_i a_i^L + a_i^L \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j = -y_i + a_i^L \sum_{j=1}^{n} y_j = -y_i + a_i^L = a_i^L - y_i$$

Доказательство $\frac{\partial C}{\partial w_{ik}^L} = a_k^{L-1}\left(a_i^L - y_i\right)$

$$\frac{\partial C}{\partial w_{ik}^L} = \frac{\partial C}{\partial z_i^L} \cdot \frac{\partial z_i^L}{\partial w_{ik}^L} = \frac{\partial C}{\partial z_i^L} \cdot a_k^{L-1} = a_k^{L-1} \frac{\partial C}{\partial z_i^L} = a_k^{L-1}\left(a_i^L - y_i\right)$$

Использование функции перекрестной энтропии в качестве функции потерь

$$C = -\sum_j y_j \ln a_j^L$$

Свойства функции потерь:

1) $C \geq 0$

2) Минимум C достигается при $\vec{a}^L = \vec{y}$:

$$\min_{\vec{a}^L} C = -\sum_j y_j \ln y_j$$

Использование функции *softmax* в качестве функции активации выходного слоя

n – количество нейронов в выходном слое L

m – количество нейронов в слое L-1

$$z_j^L = \sum_{k=1}^{m} w_{jk}^L a_k^{L-1} + b_j^L$$

Для выходного слоя L применяется особая функция активации *softmax*:

$$a_j^L = \frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}$$

Лемма 4: $\frac{\partial C}{\partial z_i^L} = a_i^L - y_i$

$$\frac{\partial C}{\partial z_i^L} = -\frac{\partial}{\partial z_i^L}\left(\sum_{j=1}^{n} y_j \ln a_j^L\right) = -\sum_{j=1}^{n} y_j \frac{\partial(\ln a_j^L)}{\partial z_i^L} = -\sum_{j=1}^{n} y_j \frac{\partial\left(\ln\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)\right)}{\partial z_i^L} = -\sum_{j=1}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L}$$

$$= -y_i \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_i^L}} \cdot \frac{\partial\left(\frac{e^{z_i^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L} - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L}$$

$$= -y_i \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_i^L}} \cdot \frac{e^{z_i^L}\sum_{i=1}^{n} e^{z_i^L} - e^{z_i^L}e^{z_i^L}}{\left(\sum_{i=1}^{n} e^{z_i^L}\right)^2} - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L} = -y_i \frac{\sum_{i=1}^{n} e^{z_i^L} - e^{z_i^L}}{\sum_{i=1}^{n} e^{z_i^L}} - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L}$$

$$= -y_i(1 - a_i^L) - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{i=1}^{n} e^{z_i^L}}\right)}{\partial z_i^L} = -y_i(1 - a_i^L) - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{\sum_{i=1}^{n} e^{z_i^L}}{e^{z_j^L}} \cdot \frac{-e^{z_j^L}e^{z_i^L}}{\left(\sum_{i=1}^{n} e^{z_i^L}\right)^2} = -y_i(1 - a_i^L) - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j \frac{-e^{z_i^L}}{\sum_{i=1}^{n} e^{z_i^L}}$$

$$= -y_i(1 - a_i^L) - \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j(-a_i^L) = -y_i(1 - a_i^L) + a_i^L \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j = -y_i + y_i a_i^L + a_i^L \sum_{\substack{j=1 \\ j \neq i}}^{n} y_j = -y_i + a_i^L \sum_{j=1}^{n} y_j = -y_i + a_i^L = a_i^L - y_i$$

Обобщение функции потерь на всю обучающую выборку

$$\mathbb{C} = -\frac{1}{|V|} \sum_{(\vec{x},y)\in V} \sum_{j} y_j \ln a_j^L$$

Леммы

Лемма 1: $\frac{\partial C}{\partial a_j^L} = -\frac{y_j}{a_j^L}$

$$\frac{\partial C}{\partial a_j^L} = -\frac{\partial}{\partial a_j^L}\left(\sum_{i=1}^{n} y_i \ln a_i^L\right) = -\frac{y_j}{a_j^L}$$

Лемма 2: $\frac{\partial a_j^L}{\partial z_j} = a_j^L(1 - a_j^L)$

$$\frac{\partial a_j^L}{\partial z_j^L} = \frac{\partial}{\partial z_j^L}\left(\frac{e^{z_j^L}}{\sum_{i=1}^n e^{z_i^L}}\right) = \frac{e^{z_j^L}\sum_{i=1}^n e^{z_i^L} - e^{2z_j^L}}{\left(\sum_{i=1}^n e^{z_i^L}\right)^2} =$$

$$= \frac{e^{z_j^L}\sum_{i=1}^n e^{z_i^L} - e^{2z_j^L}}{\left(\sum_{i=1}^n e^{z_i^L}\right)^2} = \frac{e^{z_j^L}\sum_{i=1}^n e^{z_i^L}}{\left(\sum_{i=1}^n e^{z_i^L}\right)^2} - \frac{e^{2z_j^L}}{\left(\sum_{i=1}^n e^{z_i^L}\right)^2} =$$

$$= a_j^L - \left(a_j^L\right)^2 = a_j^L(1 - a_j^L)$$

Лемма 3: $\frac{\partial a_j^L}{\partial z_k} = -a_k^L a_j^L$ при $k \neq j$

$$\frac{\partial a_j^L}{\partial z_k^L} = \frac{\partial}{\partial z_k^L}\left(\frac{e^{z_j^L}}{\sum_{i=1}^n e^{z_k^L}}\right) = \frac{-e^{z_k^L} e^{z_j^L}}{\left(\sum_{i=1}^n e^{z_i^L}\right)^2} =$$

$$= \frac{-e^{z_k^L}}{\sum_{i=1}^n e^{z_i^L}} \cdot \frac{e^{z_j^L}}{\sum_{i=1}^n e^{z_i^L}} = -a_k^L a_j^L$$

Лемма 4: $\frac{\partial C}{\partial z_i^L} = a_i^L - y_i$

$$\frac{\partial C}{\partial z_i^L} = -\frac{\partial}{\partial z_i^L}\left(\sum_{j=1}^{n} y_j \ln a_j^L\right) = -\sum_{j=1}^{n} y_j \frac{\partial\left(\ln a_j^L\right)}{\partial z_i^L} = -\sum_{j=1}^{n} y_j \frac{\partial\left(\ln\left(\frac{e^{z_j^L}}{\sum_{l=1}^{n} e^{z_l^L}}\right)\right)}{\partial z_i^L} = -\sum_{j=1}^{n} y_j \frac{\sum_{l=1}^{n} e^{z_l^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{l=1}^{n} e^{z_l^L}}\right)}{\partial z_i^L}$$

$$= -y_i \frac{\sum_{l=1}^{n} e^{z_l^L}}{e^{z_i^L}} \cdot \frac{\partial\left(\frac{e^{z_i^L}}{\sum_{l=1}^{n} e^{z_l^L}}\right)}{\partial z_i^L} - \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j \frac{\sum_{l=1}^{n} e^{z_l^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{l=1}^{n} e^{z_l^L}}\right)}{\partial z_i^L}$$

$$= -y_i \frac{\sum_{l=1}^{n} e^{z_l^L}}{e^{z_i^L}} \cdot \frac{e^{z_i^L}\sum_{l=1}^{n} e^{z_l^L} - e^{2z_i^L}}{\left(\sum_{l=1}^{n} e^{z_l^L}\right)^2} - \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j \frac{\sum_{l=1}^{n} e^{z_l^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{l=1}^{n} e^{z_l^L}}\right)}{\partial z_i^L} = -y_i \frac{\sum_{l=1}^{n} e^{z_l^L} - e^{z_i^L}}{\sum_{l=1}^{n} e^{z_l^L}} - \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j \frac{\sum_{l=1}^{n} e^{z_l^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{l=1}^{n} e^{z_l^L}}\right)}{\partial z_i^L}$$

$$= -y_i\left(1 - a_i^L\right) - \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j \frac{\sum_{l=1}^{n} e^{z_l^L}}{e^{z_j^L}} \cdot \frac{\partial\left(\frac{e^{z_j^L}}{\sum_{l=1}^{n} e^{z_l^L}}\right)}{\partial z_i^L} = -y_i\left(1 - a_i^L\right) - \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j \frac{\sum_{l=1}^{n} e^{z_l^L}}{e^{z_j^L}} \cdot \frac{-e^{z_j^L} e^{z_i^L}}{\left(\sum_{l=1}^{n} e^{z_l^L}\right)^2} = -y_i\left(1 - a_i^L\right) - \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j \frac{-e^{z_i^L}}{\sum_{l=1}^{n} e^{z_l^L}}$$

$$= -y_i\left(1 - a_i^L\right) - \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j\left(-a_i^L\right) = -y_i\left(1 - a_i^L\right) + a_i^L \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j = -y_i + y_i a_i^L + a_i^L \sum_{\substack{j=1 \\ j\neq i}}^{n} y_j = -y_i + a_i^L \sum_{j=1}^{n} y_j = -y_i + a_i^L = a_i^L - y_i$$