**Задание №18**

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$\sigma'(z) = -\frac{(1+e^{-z})'}{(1+e^{-z})^2} = -\frac{(-1)\cdot e^{-z}}{(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} =$$

$$= \sigma \cdot \frac{1+e^{-z}-1}{1+e^{-z}} = \sigma\left(1 - \frac{1}{1+e^{-z}}\right) = \sigma(1-\sigma)$$

**Задание №19**

$$g_k(s_1, s_2, ..., s_K) = \frac{e^{s_k}}{\sum_{l=1}^{K} e^{s_l}}$$

$$R^{(i)} = -\sum_{k=1}^{K} I(y^{(i)} = k) \cdot \ln g_k(s_1, s_2, ..., s_K)$$

1)

$$\frac{\partial g_k}{\partial s_l} \overset{k\neq l}{=} \frac{\partial}{\partial s_l}\left(\frac{e^{s_k}}{\sum_{i=1}^{K} e^{s_i}}\right) = \frac{-e^{s_k}e^{s_l}}{(\sum_{i=1}^{K} e^{s_i})^2} = \frac{e^{s_k}}{\sum_{i=1}^{K} e^{s_i}}\left(-\frac{e^{s_l}}{\sum_{i=1}^{K} e^{s_i}}\right) = g_k(0 - g_l)$$

$$\frac{\partial g_k}{\partial s_l} \overset{k=l}{=} \frac{\partial g_l}{\partial s_l} = \frac{\partial}{\partial s_l}\left(\frac{e^{s_l}}{\sum_{i=1}^{K} e^{s_i}}\right) = e^{s_l}\left(\sum_{i=1}^{K} e^{s_i}\right)^{-1} + e^{s_l}\left(-\frac{e^{s_l}}{(\sum_{i=1}^{K} e^{s_i})^2}\right) =$$

$$= \frac{e^{s_l}}{\sum_{i=1}^{K} e^{s_i}}\left(1 - \frac{e^{s_l}}{\sum_{i=1}^{K} e^{s_i}}\right) = g_l(1-g_l) \overset{l=k}{=} g_k(1-g_l)$$

$$\frac{\partial g_k}{\partial s_l} = g_k(I(k=l) - g_l)$$

2)

$$\frac{\partial R^{(i)}}{\partial g_k} = \frac{\partial}{\partial g_k}\left(-\sum_{k=1}^{K} I(y^{(i)} = k) \cdot \ln g_k\right) = \frac{\partial}{\partial g_k}(-I(y^{(i)}) \ln g_k) =$$

$$= -I(y^{(i)})\frac{\partial}{\partial g_k}\sum_{k=1}^{K} \ln g_k = -I(y^{(i)}) \cdot \frac{1}{g_k}$$

$$\frac{\partial R^{(i)}}{\partial g_k} = -\frac{I(y^{(i)})}{g_k}$$

3)

$$R^{(i)} = -\sum_{k=1}^{K} I(y^{(i)} = k) \ln g_k = -\sum_{k=1}^{K} I(y^{(i)} = k) \ln \left( \frac{e^{s_k}}{(\sum_{i=1}^{K} e^{s_i})^2} \right) =$$

$$= -\sum_{k=1}^{K} I(y^{(i)} = k)(\ln e^{s_k} - \ln(\sum_{i=1}^{K} e^{s_i})) = -\sum_{k=1}^{K} I(y^{(i)} = k)(s_k - \ln(\sum_{i=1}^{K} e^{s_i}))$$

$$\frac{\partial R^{(i)}}{\partial s_l} = \frac{\partial}{\partial s_l} \left( -\sum_{k=1}^{K} I(y^{(i)} = k)(s_k - \ln(\sum_{i=1}^{K} e^{s_i}) \right)$$

$$\frac{\partial R^{(i)}}{\partial s_l} \overset{k=l}{=} \frac{\partial}{\partial s_l}(-I(y^{(i)} = l)(s_l - \ln(\sum_{i=1}^{K} e^{s_i})) = -I(y^{(i)} = l) \left( 1 - \frac{e^{s_l}}{\sum_{i=1}^{K} e^{s_i}} \right) =$$

$$= -I(y^{(i)} = l)(1 - g_l) = -I(y^{(i)} = l) + I(y^{(i)} = l) \cdot g_l$$

$$\frac{\partial R^{(i)}}{\partial s_l} \overset{k \neq l}{=} \frac{\partial}{\partial s_l}(-I(y^{(i)} = k)(s_k - \ln(\sum_{i=1}^{K} e^{s_i})) = -I(y^{(i)} = k) \left( -\frac{e^{s_l}}{\sum_{i=1}^{K} e^{s_i}} \right) =$$

$$= I(y^{(i)} = l) \cdot g_l$$

$$\frac{\partial R^{(i)}}{\partial s_l} = \left( \sum_{k=1}^{K} I(y^{(i)} = k) \right) g_l - I(y^{(i)} = l) = g_l - I(y^{(i)} = l)$$

**Задание №21**

$R^{(i)} = \mathrm{logloss}(\mathrm{softmax}(B(\sigma(Ax))))$

Обозначим функцию logloss как $L$, функцию softmax как $g$. Тогда:

$R^{(i)} = L(g(B \cdot \sigma(Ax)))$

$v(x) = B \cdot \sigma(Ax)$

$w(x) = Ax$

$R^{(i)} = L(g(v(x)))$

$$\frac{\partial R^{(i)}}{\partial x} = \frac{\partial L}{\partial g}\frac{\partial g}{\partial x} = \frac{\partial L}{\partial g}\frac{\partial g}{\partial v}\frac{\partial v}{\partial x} = \frac{\partial L}{\partial v}\frac{\partial v}{\partial x} = \frac{\partial L}{\partial v}\frac{\partial (B \cdot \sigma(w))}{\partial x} = \frac{\partial L}{\partial v}\frac{\partial (B \cdot \sigma(w))}{\partial \sigma}\frac{\partial \sigma}{\partial x} =$$

$$= \frac{\partial R^{(i)}}{\partial v}\frac{\partial (B \cdot \sigma(Ax))}{\partial \sigma(Ax)}\frac{\partial \sigma(Ax)}{\partial Ax}\frac{\partial Ax}{\partial x}$$

$$\frac{\partial R^{(i)}}{\partial v} = g - y, \frac{\partial (B \cdot \sigma(Ax))}{\partial \sigma(Ax)} = B, \frac{\partial \sigma(Ax)}{\partial Ax} = \mathrm{diag}(\sigma'), \frac{\partial Ax}{\partial x} = A$$

$$\frac{\partial R^{(i)}}{\partial x} = (g - y) \cdot B \cdot \mathrm{diag}(\sigma') \cdot A$$

$$\frac{\partial R^{(i)}}{\partial A} = (g - y) \cdot B \cdot \mathrm{diag}(\sigma') \cdot x$$

$$\frac{\partial R^{(i)}}{\partial B} = (g - y) \cdot \sigma(Ax)$$