Deep Learning

3.1.1. Please, write down sizes of the following matrices and vectors where N is a batch:

(a) $\mathbf{w}^{(in)}$ - the matrix 75×225
$\mathbf{b}^{(in)}$ - the vector 75×1
$\mathbf{w}^{(rec)}$ - the matrix 75×75
$\mathbf{b}^{(rec)}$ - the vector 75×1
$\mathbf{w}^{(link)}$ - the matrix 75×75
$\mathbf{w}^{(out)}$ - the matrix 225×75
$\mathbf{b}^{(out)}$ - the vector 225×1

(b) $\mathbf{x}$ - the vector 225×1
$\mathbf{x}^{(reduce)}$ - the vector 75×1
$\mathbf{a}^{(in)}$ - the vector 75×1
$\mathbf{z}^{(in)}$ - the vector 75×1
$\mathbf{a}^{(rec)}$ - the vector 75×1
$\mathbf{z}^{(rec)}$ - the vector 75×1
$\mathbf{a}^{(link)}$ - the vector 75×1
$\mathbf{z}^{(link)}$ - the vector 75×1
$\mathbf{z}^{(rec+link)}$ - the vector 75×1
$\mathbf{a}^{(out)}$ - the vector 225×1
$\mathbf{y}$ - the vector of 225×1

3.1.2. Please, write down forward equations for $Layer_{link}$, $Layer_{rec}$, $Layer_{out}$ in a scalar form.
Solution.
$Layer_{link}$ equation:

$$a_{n,j}^{(link)} = \sum_{i=1}^{D} w_{j,i}^{(link)} * x_{n,i} + b_j^{(link)},$$

where $1 \leq n \leq N_{batch}$, $1 \leq i \leq D = 225/3 = 75$, $1 \leq j \leq P = 75$

$Layer_{rec}$ equation: There are two cyles in this layer. $\mathbf{z}^{(in)}$ is an input in the first iteration. So,

$$a_{n,j}^{(rec)} = \sum_{i=1}^{D} w_{j,i}^{(rec)} * z_{n,i}^{(in)} + b_j^{(rec)},$$

where $1 \leq n \leq N_{batch}$, $1 \leq i \leq D = 75$, $1 \leq j \leq P = 75$. Output after the first cycle will be:

$$z_{n,j}^{(rec)} = ReLU(a_{n,j}^{(rec)})$$

In the second cycle $d^{-1} = z_{n,j}^{(rec)}$ will be an input for the same layer plus $z_{n,i}^{(in)}$ and:

$$a_{n,j}^{(rec)} = \sum_{i=1}^{D} w_{j,i}^{(rec)} * (z_{n,i}^{(in)} + d_{n,i}^{-1}) + b_j^{(rec)},$$

$$z_{n,j}^{(rec)} = ReLU(a_{n,j}^{(rec)})$$

where $1 \leq n \leq N_{batch}$, $1 \leq i \leq D = 75$, $1 \leq j \leq P = 75$.

$Layer_{out}$ equation: Input for this layer will be the sum of outputs from $Layer_{rec}$ and $Layer_{link}$

$$z_{n,j}^{(rec+link)} = z_{n,j}^{(rec)} + z_{n,j}^{(link)}$$

$$a_{n,j}^{(out)} = \sum_{i=1}^{D} w_{j,i}^{(out)} * z_{n,j}^{(rec+link)} + b_j^{(out)},$$

where $1 \leq n \leq N_{batch}$, $1 \leq i \leq D = 75$, $1 \leq j \leq P = 225$

3.1.3. Please, write down forward equations for $Layer_{in}$, $Layer_{link}$, $Layer_{out}$, $Layer_{rec}$ in a vector form.

Solution.
$Layer_{(in)}$ equation:

$$A_{in} = X * W_{in}^{\top} + b_{in}$$

$$Z_{in} = ReLU(A_{in})$$

where X - input batch $225 \times N$, W - weight matrix, b - bias values

$Layer_{(link)}$ equation: X_{reduce} is equal to input matrix X multiplied by matrix K of size n rows = 225 and m columns = 75. K matrix is filled in with zeroes except $K_{(row,column)} = K_{(3n-2,n)} = 1$, where $1 \leq n \leq 74$

$$A_{link} = X * K * W_{link}^{\top}$$

$$Z_{link} = ReLU(A_{link})$$

where X_{reduce} - reduced matrix from input X , W_{link} - weight matrix

$Layer_{(rec)}$ equation: In the first cycle Z_{in} will be input from the $Layer_{(in)}$. d^{-1} will be 0 in the first cycle. So,

$$A_{rec(1)} = (Z_{in} + 0) * W_{rec}^{\top} + b_{rec}$$

$$Z_{rec(1)} = ReLU(A_{rec(1)})$$

After the second cycle

$$A_{rec(2)} = (Z_{in} + Z_{rec(1)}) * W_{rec}^{\top} + b_{rec}$$

$$Z_{rec(2)} = ReLU(A_{rec(2)})$$

$Layer_{(out)}$ equation: Input for this layer will be the sum of outputs from $Layer_{(rec)}$ and $Layer_{(link)}$.

$$Z_{rec+link} = Z_{rec(2)} + Z_{link}$$

$$A_{out} = Z_{rec+link} * W_{out}^{\top} + b_{out}$$
$$y = ReLU(A_{in})$$

3.2.1. Please, write down backward pass flow for local gradients (AKA "deltas") δ for each layer in scalar form.
3.2.2. Please, write down the following derivatives for neural network's weights updates.
Solution. The sum squared error formula is:

$$E(w) = \sum_{n=1}^{N} (x_n - y_n(x_n w))^2$$

3.2.1-3.2.2 a) We start the back propagation path from the last layer, so derivative of error function with respect to w in $Layer_{out}$ will be:

$$\frac{\mathrm{d}E}{\mathrm{d}w_{k,j}^{(out)}} = 2\sum_{n=1}^{N}(x_n - y_n)\frac{\mathrm{d}y_n}{\mathrm{d}w_{k,j}^{(out)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{k,j}^{(out)}} = 2\sum_{n=1}^{N}(x_n - y_n)\frac{\mathrm{d}f(a_n^{(out)})}{\mathrm{d}a_n^{(out)}}\frac{\mathrm{d}a_n^{(out)}}{\mathrm{d}w_{k,j}^{(out)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{k,j}^{(out)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})z_j^{(rec+link)}$$

$$\delta_n^{out} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{k,j}^{(out)}} = \delta_n^{out} z_j^{(rec+link)}$$

3.2.2.a. Derivative with respect to b:

$$\frac{\mathrm{d}E}{\mathrm{d}b^{(out)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)}) * 1$$

b) Derivatives of error function with respect to w in $Layer_{link}$:

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(link)}} = 2\sum_{n=1}^{N}(x_n - y_n)\frac{\mathrm{d}y_n}{\mathrm{d}w_{j,i}^{(link)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(link)}} = 2\sum_{n=1}^{N}(x_n - y_n)\frac{\mathrm{d}f(a_n^{(out)})}{\mathrm{d}a_n^{(out)}}\frac{\mathrm{d}a_n^{(out)}}{\mathrm{d}w_{j,i}^{(link)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(link)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}\frac{\mathrm{d}(z_j^{(link)} + z_j^{(rec)})}{\mathrm{d}w_{j,i}^{(link)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(link)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}x_{reduce}$$

$$\delta_j^{link} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(link)}} = \delta_j^{link}x_{reduce}$$

c) Derivatives of error function with respect to w in $Layer_{rec}$:

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(rec)}} = 2\sum_{n=1}^{N}(x_n - y_n)\frac{\mathrm{d}y_n}{\mathrm{d}w_{j,i}^{(rec)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(rec)}} = 2\sum_{n=1}^{N}(x_n - y_n)\frac{\mathrm{d}f(a_n^{(out)})}{\mathrm{d}a_n^{(out)}}\frac{\mathrm{d}a_n^{(out)}}{\mathrm{d}w_{j,i}^{(rec)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(rec)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}\frac{\mathrm{d}(z_j^{(link)} + z_j^{(rec)})}{\mathrm{d}w_{j,i}^{(rec)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(rec)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}\frac{\mathrm{d}(z_j^{(rec)})}{\mathrm{d}w_{j,i}^{(rec)}}$$

$$a_j^{(rec*)} - the\ first\ cycle\ output,\ a_j^{(rec**)} - the\ second\ cycle\ output,$$

$$\frac{\mathrm{d}z_j^{(rec)}}{\mathrm{d}w_{j,i}^{(rec)}} = f'(a_j^{(rec**)})\frac{\mathrm{d}a_j^{(rec**)}}{\mathrm{d}w_{j,i}^{(rec)}}$$

$$\delta_n^{rec**} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}f'(a_j^{(rec**)})$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(rec)}} = \delta_n^{rec**} * \frac{\mathrm{d}a_j^{(rec**)}}{\mathrm{d}w_{j,i}^{(rec)}}$$

$$\frac{\mathrm{d}a_j^{(rec**)}}{\mathrm{d}w_{j,i}^{(rec)}} = \frac{\mathrm{d}f(a_j^{(rec**)})}{\mathrm{d}a_j^{(rec**)}}\frac{\mathrm{d}a_j^{(rec**)}}{\mathrm{d}w_{j,i}^{(rec)}}\frac{\mathrm{d}f(a_j^{(rec*)})}{\mathrm{d}a_j^{(rec*)}}\frac{\mathrm{d}a_j^{(rec*)}}{\mathrm{d}w_{j,i}^{(rec)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{j,i}^{(rec)}} = \delta_n^{rec**}\frac{\mathrm{d}f(a_j^{(rec**)})}{\mathrm{d}a_j^{(rec**)}}\frac{\mathrm{d}a_j^{(rec**)}}{\mathrm{d}w_{j,i}^{(rec)}}\frac{\mathrm{d}f(a_j^{(rec*)})}{\mathrm{d}a_j^{(rec*)}}\frac{\mathrm{d}a_j^{(rec*)}}{\mathrm{d}w_{j,i}^{(rec)}}$$

$$3.2.1 - 3.2.\ d)\ Derivatives\ of\ error\ function\ in\ (Layer_{in}):$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{i,r}^{(in)}} = 2\sum_{n=1}^{N}(x_n - y_n)\frac{\mathrm{d}y_n}{\mathrm{d}w_{i,r}^{(in)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{i,r}^{(in)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}\frac{\mathrm{d}(z_j^{(rec)})}{\mathrm{d}w_{i,r}^{(in)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{i,r}^{(in)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}f'(a_j^{(rec**)})\frac{\mathrm{d}(a_j^{(rec**)})}{\mathrm{d}w_{i,r}^{(in)}}$$

$$\frac{\mathrm{d}E}{\mathrm{d}w_{i,r}^{(in)}} = 2\sum_{n=1}^{N}(x_n - y_n)f'(a_n^{(out)})w_{n,j}^{(out)}f'(a_j^{(rec**)})\frac{\mathrm{d}f(a_j^{(rec**)})}{\mathrm{d}a_j^{(rec**)}}\frac{\mathrm{d}a_j^{(rec**)}}{\mathrm{d}w_{i,r}^{(rec)}}\frac{\mathrm{d}f(a_j^{(rec*)})}{\mathrm{d}a_j^{(rec*)}}\frac{\mathrm{d}a_j^{(rec*)}}{\mathrm{d}w_{i,r}^{(in)}} *$$

$$\frac{\mathrm{d}f(a_k^{(in)})}{\mathrm{d}a_j^{(in)}}\frac{\mathrm{d}a_k^{(in)}}{\mathrm{d}w_{i,r}^{(in)}}$$