

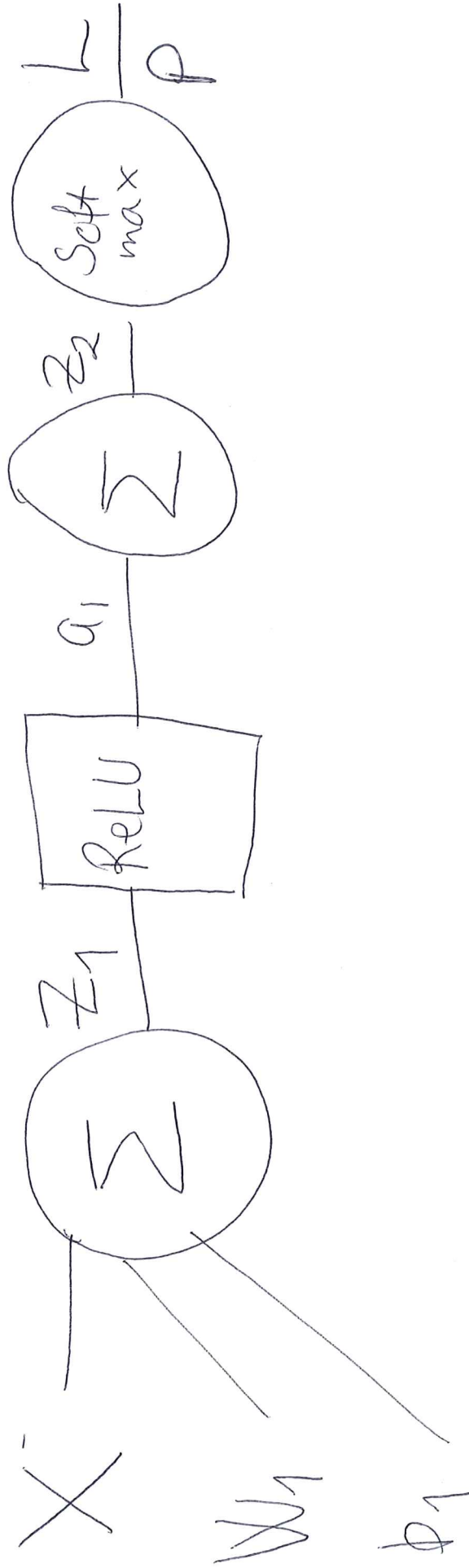
$$X = 200 \times 4$$

$$W_1 = 4 \times 10$$

$$b_1 = 1 \times 10$$

$$W_2 = 10 \times 3$$

$$b_2 = 1 \times 3$$



$$a_1 = 200 \times 10$$

Do step-by-step backpropagation back to $\frac{\partial L}{\partial W_2}$, $\frac{\partial L}{\partial b_2}$, $\frac{\partial L}{\partial W_1}$, $\frac{\partial L}{\partial b_1}$

Create Indicators matrix Y_{ind} , $Y_{ind}[i, k] = 1$ if x_i is from class k

$$Y_{ind} = 200 \times 3$$

$$p = np.exp(scores)$$

$$p /= p.sum(axis=1, keepdims=True)$$

$$\frac{\partial L}{\partial Z_2}$$

P: 200x3-matrix

P = matrix of softmax probabilities for 3 classes, N=200

$$P[i,:] = [p_i(\text{klasse1}) \quad p_i(\text{klasse2}) \quad p_i(\text{klasse3})]$$

Create indicator matrix Y_{ind}

belongs to class k

$$Y_{ind}[i,k] = 1 \text{ if sample } Y_i \text{ belongs to class } k$$

Compute $\frac{\partial L}{\partial Z_2}$ vectorized for all samples

$$\partial Z_2 \text{ mat} = (p - Y_{ind}) | N$$

$$\partial Z_2 \text{ mat shape} = 200 \times 3$$

$P = np.exp(scores)$
 $P = p.sum(axis=1, keepdim=True)$

$$\frac{\partial L}{\partial W_2} = (p - y_{ind}) \cdot a_1$$

$200 \times 3 \quad 200 \times 10$

~~dzvec~~
 dzvec-variable

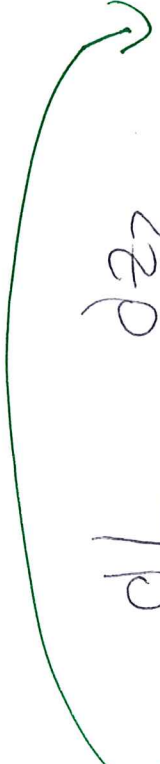
$W_2: 10 \times 3$ want 10×3 :

$$z_2 = W_2 \cdot a_1 + b_2 \quad \frac{\partial L}{\partial W_2} = a_1.T.dot(p - y_{ind}) + reg * W_2$$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial b_2} = dz_2vec.sum(0)$$

~~z1~~

$$\frac{\partial L}{\partial a_1}$$



$$= \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial a_1}$$

$$da_{1mat} = \partial z_2 vec : dot (W_2.T)$$

$$= \frac{\partial L}{\partial a_1} \cdot \frac{\partial a_1}{\partial w_1} = da_{1mat}$$

$$da_{1mat} [a_1 = 0] = 0$$

Backwards through ReLU from a_1 to z_1

$$\frac{\partial a_1}{\partial w_1} = X \text{ if } a_1 > 0, \text{ 0 otherwise}$$

$$\frac{\partial L}{\partial b_1} = \partial da_{1mat}.sum(0)$$

$$\frac{\partial L}{\partial w_1} = X.T.dot(da_{1mat}) + reg.w_1$$