CV-Assignment9

vezeteu eugeniu - 886240

November 2020

1 Task

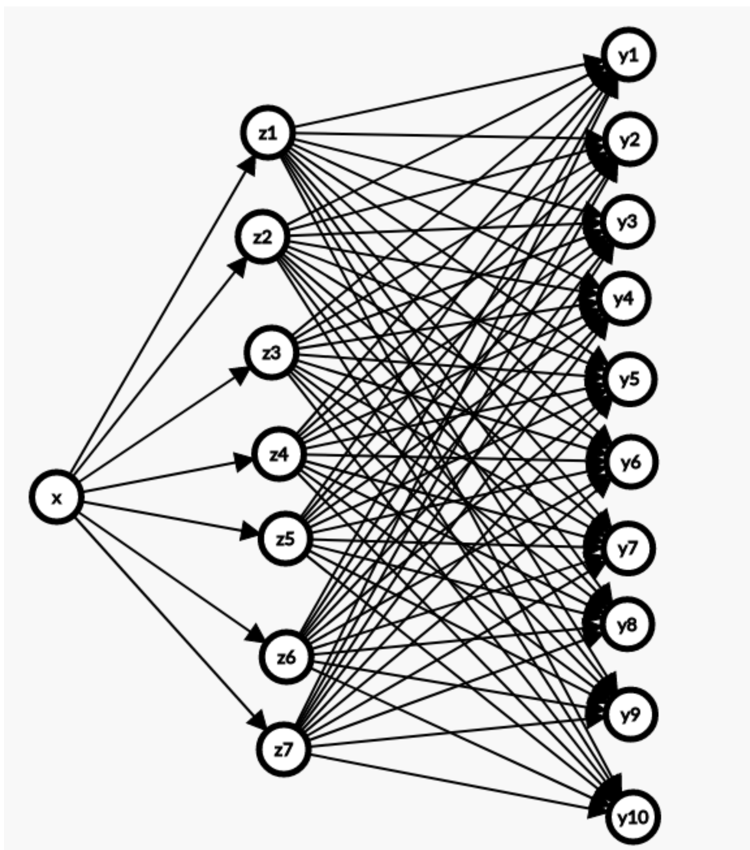

Figure 1: Network arhitecture

1. Equation 1 is $E = \frac{1}{m}\sum_{j=1}^{m} -t_j \cdot log(y_i)$

With m=1 we have

$$E = -t * log(y) = -t * log(sigmoid(Wx))$$

$$= -t * log(\frac{1}{1 + e^{-W*x}})$$

Taking derivative of E w.r.t. t resuts:

$$\frac{\partial E}{\partial t} = -log(sigmoid(Wx)) = -log(\frac{1}{1 + e^{-W*x}})$$

Taking derivative of E w.r.t. x resuts:

$$\frac{\partial E}{\partial x} = -t * log(sigmoid(Wx))' * \frac{\partial sigmoid(Wx)}{\partial x} =$$

$$\frac{\partial E}{\partial x} = -t * \frac{1}{sigmoid(Wx)} * sigmoid\_derivative(Wx) * W =$$

$$\frac{\partial E}{\partial x} = -t * (1 + e^{-Wx}) * \frac{1}{1 + e^{-W*x}} * (1 - \frac{1}{1 + e^{-W*x}}) * W =$$

$$\frac{\partial E}{\partial x} = \frac{-t * W * e^{e^{-Wx}}}{1 + e^{-Wx}}$$

2. Using chain rule we have: $\frac{\partial E}{\partial z} = \frac{\partial E}{\partial y} * \frac{\partial y}{\partial z}$ We must show that: $\frac{W}{\partial z^{(2)}} = (y^{(2)} - t)^T$ Taking the derivative of softmax results:

$$\frac{\partial y}{\partial z} = \begin{cases} \frac{\partial softmax(z)}{\partial z} = y * (1 - y) \quad for \quad i = j \\ \frac{\partial softmax(z)}{\partial z} = -y_i * y_j \quad for \quad i \neq j \end{cases}$$

$$\frac{\partial E}{\partial z^{(2)}} = \sum_{i=j} \frac{\partial E}{\partial y_i} * \frac{\partial y_i}{\partial z_i} + \sum_{i \neq j} \frac{\partial E}{\partial y_j} * \frac{\partial y_j}{\partial z_i}$$

which is:

$$\frac{\partial E}{\partial z^{(2)}} = \sum_{i=j} \frac{-t_j}{y_i} * y_i * (1 - y_i) + \sum_{i \neq j} \frac{-t_j}{y_i} * -y_i * y_j$$

$$\frac{\partial E}{\partial z^{(2)}} = -t * (1 - y) + t * y$$

Assume that elements if t_j sum to 1, we have:

$$\frac{\partial E}{\partial z^{(2)}} = -t + y = y - t$$

So, $\frac{W}{\partial z^{(2)}} = (y^{(2)} - t)^T$

3. Show that $\frac{\partial E}{\partial y^{(1)}} = (y^{(2)} - t)^T * W^{(2)}$

$$\frac{\partial E}{\partial y^{(1)}} = \frac{\partial E}{\partial z^{(2)}} * \frac{\partial z^{(2)}}{\partial y^{(1)}} = (y^{(2)} - t)^T * W^{(2)}$$

4.

$$\frac{\partial E}{\partial w_{uv}^{(2)}} = \frac{\partial E}{\partial z^{(2)}} * \frac{\partial z^{(2)}}{\partial w_{uv}^{(2)}} = (y^{(2)} - t) * y^{(1)T}$$

5. Derivative of sigmoid function is:

$$\frac{\partial sigmoid(z)}{\partial z} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} * \frac{e^{-z}}{1+e^{-z}} = sigmoid(z) * (1 - sigmoid(z))$$

Therefore,

$$\frac{\partial y^{(1)}}{\partial z^{(1)}} = diag(y^{(1)} . * (1 - y^{(1)}))$$

6. Show that $\frac{\partial E}{\partial z^{(1)}} = (y^{(2)} - t)^T W^{(2)} diag(y^{(1)} . * (1 - y^{(1)}))$ Again, using chain rule

$$\frac{\partial E}{\partial z^{(1)}} = \frac{\partial E}{\partial y^{(1)}} * \frac{\partial y^{(1)}}{\partial z^{(1)}} = (y^{(2)} - t)^T W^{(2)} diag(y^{(1)} . * (1 - y^{(1)}))$$

7.

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial z^{(1)}} * \frac{\partial z^{(1)}}{\partial W^{(1)}} = \frac{\partial E}{\partial z^{(1)}} * \frac{\partial (W^{(1)} * x)}{\partial W^{(1)}} = \frac{\partial E}{\partial z^{(1)}} * x$$

8. If we have m¿1, we compute above for each batch of data.

9. In case of L_2 normalization, we normalize the partial derivatives also, by multiplying with λ and value of the W. $(\lambda * W)$