House prices

1) Implement by hand equations of derivative using LaTeX (I will use similar format in LibreOffice/ODF, sorry)

$$f_x=\frac{\partial f}{\partial x}=\lim_{h\to 0}\frac{f(x+h,y)-f(x,y)}{h}$$

$$f_x=\frac{\partial f}{\partial y}=\lim_{h\to 0}\frac{f(x,y+h)-f(x,y)}{h}$$

why did I wrote them here?

2) Writing equations for housings:

$$dw\,linear:\frac{\partial}{\partial W}[W\cdot x+b]=x$$

$$dx\,linear:\frac{\partial}{\partial x}[W\cdot x+b]=W$$

$$db\,linear:\frac{\partial}{\partial b}[W\cdot x+b]=1$$

3) Cost function
We will use mean squared error cost function:

$$J(\theta_0,\theta_1)=L_{MSE}=\frac{1}{N}\cdot\sum_{i=0}^{N}(h_\theta(x_i)-y_i)^2$$

where $h_\theta(x_i)=\theta_0+\theta_1\cdot x_i$ (is our model)
We need to minimize cost function’s result.

4) Gradient descent
The goal of this is to update θ_0, b and θ_1, W to minimize cost function result.

$$\theta_0:=\theta_0-\alpha\cdot\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1)$$

$$b:=b-\alpha\cdot\frac{\partial}{\partial b}J(b,W)$$

$$\theta_1:=\theta_1-\alpha\cdot\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1)$$

$$W:=W-\alpha\cdot\frac{\partial}{\partial W}J(b,W)$$

but to simultaneously update both variables we need to assign them to temporary variables first, so b doesn’t affect calculation of the W.

α - is a learning rate, which defines how fast we are going to change b and W.

By intuition what is going to happen – partial derivative of the cost function will be a positive slope (will return a positive number) if the mse will on the right side of the local minima, as a result we will subtract a slope value multiplied by learning rate from W (or b). Otherwise we will add it. In ideal situation once MSE is 0 – we don’t move anymore.

So let’s try to figure out partial derivatives:
For θ_0 or b :

$$\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1)=\frac{\partial}{\partial\theta_0}\cdot\frac{1}{N}\cdot\sum_{i=0}^{N}(h_\theta(x_i)-y_i)^2=\frac{\partial}{\partial\theta_0}\cdot\frac{1}{N}\cdot\sum_{i=0}^{N}(\theta_0+\theta_1x_i-y_i)^2=\frac{2}{N}\sum_{i=0}^{N}(\theta_0+\theta_1x_i-y_i)\cdot 1$$

equals to:

$$\frac{2}{N}\sum_{i=0}^{N}(\theta_0+\theta_1 x_i-y_i)\cdot 1=\frac{2}{N}\sum_{i=0}^{N}(b+W\cdot x_i-y_i)\cdot 1$$

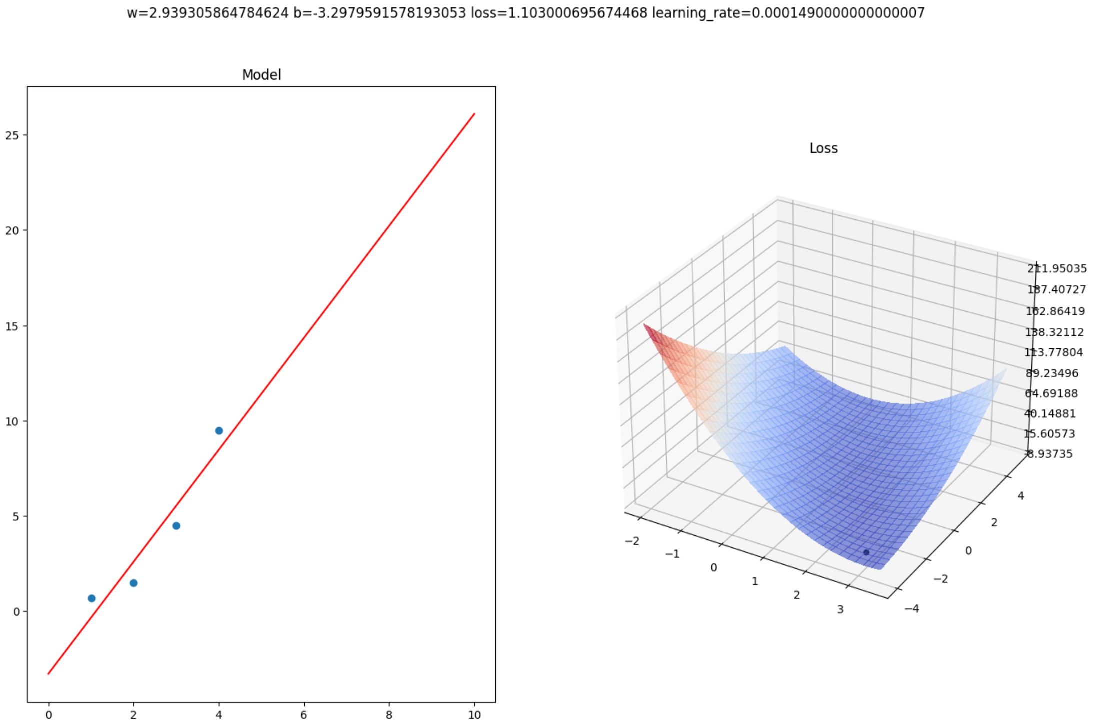
and for θ_1 or W :

$$\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1)=\frac{\partial}{\partial\theta_1}\cdot\frac{1}{N}\cdot\sum_{i=0}^{N}(h_\theta(x_i)-y_i)^2=\frac{\partial}{\partial\theta_1}\cdot\frac{1}{N}\cdot\sum_{i=0}^{N}(\theta_0+\theta_1 x_i-y_i)^2=\frac{2}{N}\sum_{i=0}^{N}(\theta_0+\theta_1 x_i-y_i)\cdot x_i$$

equals to:

$$\frac{2}{N}\sum_{i=0}^{N}(\theta_0+\theta_1 x_i-y_i)\cdot x_i=\frac{2}{N}\sum_{i=0}^{N}(b+W\cdot x_i-y_i)\cdot x_i$$

For linear model:



For sigmoid model:

$$\frac{1}{1+e^{-x}}$$

We need to find a gradient descent:

$$\theta_0:=\theta_0-\alpha\cdot\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1)$$

$$b:=b-\alpha\cdot\frac{\partial}{\partial b}J(b,W)$$

$$\theta_1:=\theta_1-\alpha\cdot\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1)$$

$$W:=W-\alpha\cdot\frac{\partial}{\partial W}J(b,W)$$

$$\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1)=\frac{\partial}{\partial\theta_0}\cdot\frac{1}{N}\cdot\sum_{i=0}^{N}(h_\theta(x_i)-y_i)^2$$

where $h_\theta(x)=\frac{1}{1+e^{-(\theta_0+\theta_1\cdot x)}}=\frac{1}{1+e^{-(b+W\cdot x)}}$

let’s substitute $a=b+W\cdot x$ so $h_\theta(x)=\frac{1}{1+e^{-a}}$

$$\frac{\partial}{\partial a}\frac{1}{1+e^{-a}}=\frac{\partial}{\partial a}(1+e^{-a})^{-1}=-(1+e^{-a})^{-2}\cdot\frac{\partial}{\partial a}(1+e^{-a})=-(1+e^{-a})^{-2}\cdot(\frac{\partial}{\partial a}1+\frac{\partial}{\partial a}e^{-a})=$$

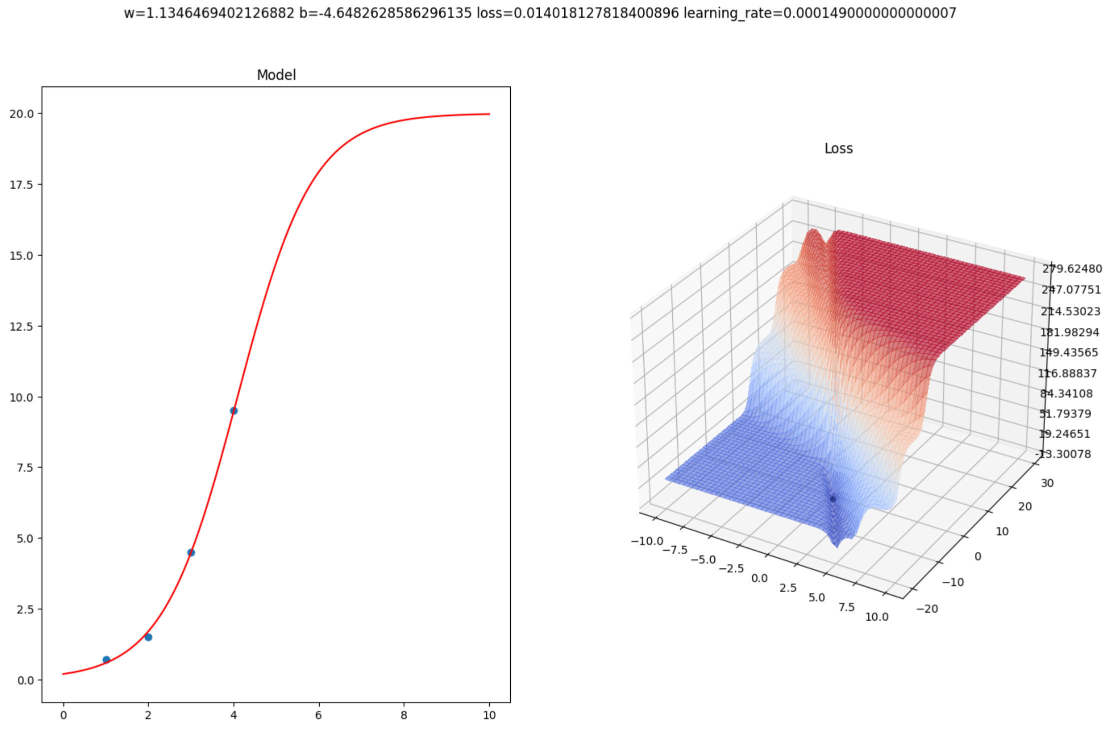
$$=-(1+e^{-a})^{-2}\cdot(0+e^{-a}\cdot\frac{\partial}{\partial a}[-a])=-(1+e^{-a})^{-2}\cdot(e^{-a}\cdot-1)=\frac{e^{-a}}{(1+e^{-a})^2}=$$

$$=\frac{e^{-a}}{(1+e^{-a})\cdot(1+e^{-a})}=\frac{1\cdot e^{-a}}{(1+e^{-a})\cdot(1+e^{-a})}=\frac{1}{1+e^{-a}}\cdot\frac{e^{-a}}{1+e^{-a}}=$$

$$=\frac{1}{1+e^{-a}}\cdot\frac{e^{-a}+1-1}{1+e^{-a}}=\frac{1}{1+e^{-a}}\cdot(\frac{1+e^{-a}}{1+e^{-a}}-\frac{1}{1+e^{-a}})=\frac{1}{1+e^{-a}}\cdot(1-\frac{1}{1+e^{-a}})$$

$$\frac{\partial}{\partial\theta_0}\frac{1}{1+e^{-a}}=\frac{\partial}{\partial b}\frac{1}{1+e^{-a}}=\frac{\partial}{\partial a}\cdot\frac{\partial}{\partial b}=\frac{e^{-a}}{(1+e^{-a})^2}\cdot 1=\frac{1}{1+e^{-a}}\cdot(1-\frac{1}{1+e^{-a}})\cdot 1$$

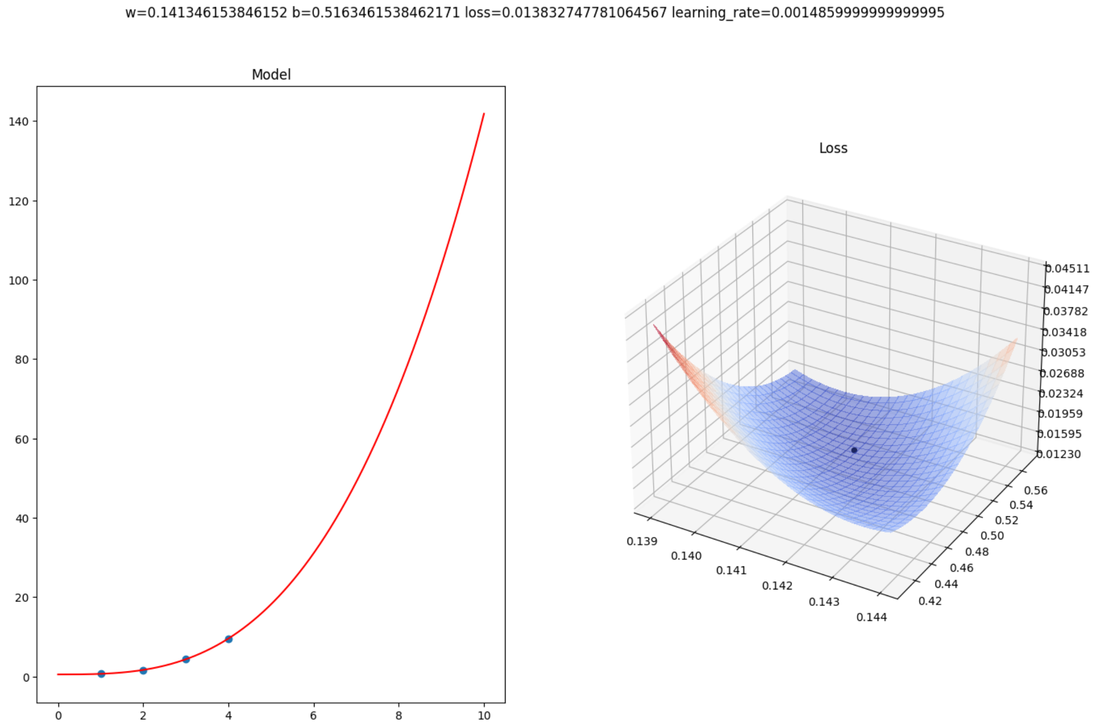
$$\frac{\partial}{\partial\theta_1}\frac{1}{1+e^{-a}}=\frac{\partial}{\partial W}\frac{1}{1+e^{-a}}=\frac{\partial}{\partial a}\cdot\frac{\partial}{\partial W}=\frac{e^{-a}}{(1+e^{-a})^2}\cdot x=\frac{1}{1+e^{-a}}\cdot(1-\frac{1}{1+e^{-a}})\cdot x$$



Animated version: https://www.youtube.com/watch?v=4hFCo9tbU34

It looks cool but I don’t find sigmoid logical here because there is no chance that 10 floor building may cost same as 100 floor. What about cubic $b+W\cdot x^3$? Where b would shift the line vertically and W would regulate its width. That was my intuition on how to fit the line. The nice part here is that loss function derivative dW, db is the same as for linear. And dx is:

$$dx\,cubic:\frac{\partial}{\partial x}[W\cdot x^3+b]=W\cdot 3\cdot x^2$$



And it worked. That’s much better.

Backpropagation

$$y' = M(x) = Linear(W_1, b_1, W_2, b_2, x) = Linear(W_2, b_2, ReLU(Linear(W_1, b_1, x)))$$

$$Linear(W_i, b_i, x_i) = W_i \cdot x_i + b_i$$

$$ReLU(x_i) = \begin{cases} x_i, x_i \geq 0 \\ 0, x_i < 0 \end{cases}$$

$$MAE(y', y) = \frac{1}{N} \sum_{i=1}^{N} (y_i - y_i')$$

SGD:

$$W_i' = W_i - \alpha \cdot \frac{MAE(y, W_1, b_1, W_2, b_2, x)}{\partial W_i}$$

$$b_i' = b_i - \alpha \cdot \frac{MAE(y, W_1, b_1, W_2, b_2, x)}{\partial b_i}$$

$$\frac{MAE(y, M(x))}{\partial W_i} = \frac{|y - M(x)|}{\partial W_i}$$

Let's assume: $a = y - M(x)$

Then: $$\frac{|a|}{\partial a} = \frac{\sqrt{a^2}}{\partial a} = \frac{(a^2)^{\frac{1}{2}}}{\partial a} = \frac{1}{2} \cdot (a^2)^{-\frac{1}{2}} \cdot \frac{a^2}{\partial a} = \frac{1}{2} \cdot (a^2)^{-\frac{1}{2}} \cdot 2a = a \cdot (a^2)^{-\frac{1}{2}} = a \cdot \frac{1}{|a|} = \frac{a}{|a|} = \frac{y - M(x)}{|y - M(x)|}$$

$$\frac{Linear(W_i, b_i, x)}{\partial W_i} = \frac{W_i \cdot x + b_i}{\partial W_i} = x$$

$$\frac{Linear(W_i, b_i, x)}{\partial b_i} = \frac{W_i \cdot x + b_i}{\partial b_i} = 1$$

$$\frac{MAE(y, W_1, b_1, W_2, b_2, x)}{\partial W_2} = \frac{|a|}{\partial a} \cdot \frac{M(x)}{\partial W_2} = \frac{y - M(x)}{|y - M(x)|} \cdot \frac{Linear(W_2, b_2, ReLU(Linear(W_1, b_1, x)))}{\partial W_2} =$$

$$= \frac{y - M(x)}{|y - M(x)|} \cdot ReLU(Linear(W_1, b_1, x))$$

$$\frac{MAE(y, W_1, b_1, W_2, b_2, x)}{\partial b_2} = \frac{|a|}{\partial a} \cdot \frac{M(x)}{\partial b_2} = \frac{y - M(x)}{|y - M(x)|} \cdot \frac{Linear(W_2, b_2, ReLU(Linear(W_1, b_1, x)))}{\partial b_2} =$$

$$= \frac{y - M(x)}{|y - M(x)|} \cdot 1$$

$$M(x) = Linear(W_2, b_2, ReLU(Linear(W_1, b_1, x)))$$

$$z = ReLU(q)$$

$$q = Linear(W_1, b_1, x)$$

$$M(x) = Linear(W_2, b_2, z) = W_2 \cdot z + b_2$$

$$\frac{MAE(y,W_1,b_1,W_2,b_2,x)}{\partial W_1}=\frac{|a|}{\partial a}\cdot\frac{M(x)}{\partial W_1}=\frac{|a|}{\partial a}\cdot\frac{Linear(W_2,b_2,z)}{\partial W_1}=\frac{|a|}{\partial a}\cdot\frac{Linear(W_2,b_2,z)}{\partial z}\cdot\frac{z}{\partial W_1}=$$

$$=\frac{|a|}{\partial a}\cdot\frac{W_2\cdot z+b_2}{\partial z}\cdot\frac{z}{\partial W_1}=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(Linear(W_1,b_1,x))}{\partial W_1}=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot\frac{q}{\partial W_1}=$$

$$=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot\frac{Linear(W_1,b_1,x)}{\partial W_1}=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot\frac{W_1\cdot x+b_1}{\partial W_1}=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot x=$$

$$=\frac{y-M(x)}{|y-M(x)|}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot x$$

$$\frac{ReLU(q)}{\partial q}=\begin{cases}1,q\geq 0\\0,q<0\end{cases}$$

$$\frac{ReLU(W_1\cdot x+b_1)}{\partial[W_1\cdot x+b_1]}=\begin{cases}1,W_1\cdot x+b_1\geq 0\\0,W_1\cdot x+b_1<0\end{cases}$$

$$\frac{MAE(y,W_1,b_1,W_2,b_2,x)}{\partial b_1}=\frac{|a|}{\partial a}\cdot\frac{M(x)}{\partial b_1}=\frac{|a|}{\partial a}\cdot\frac{Linear(W_2,b_2,z)}{\partial b_1}=\frac{|a|}{\partial a}\cdot\frac{Linear(W_2,b_2,z)}{\partial z}\cdot\frac{z}{\partial b_1}=$$

$$=\frac{|a|}{\partial a}\cdot\frac{W_2\cdot z+b_2}{\partial z}\cdot\frac{z}{\partial b_1}=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(Linear(W_1,b_1,x))}{\partial b_1}=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot\frac{q}{\partial b_1}=$$

$$=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot\frac{Linear(W_1,b_1,x)}{\partial b_1}=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot\frac{W_1\cdot x+b_1}{\partial b_1}=\frac{|a|}{\partial a}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}\cdot 1=$$

$$=\frac{y-M(x)}{|y-M(x)|}\cdot W_2\cdot\frac{ReLU(q)}{\partial q}$$

Model: $y'=M(x)=LeakyReLU(Linear(\tanh(Linear(W\cdot x+b))))$

$$y'=M(x)=LeakyReLu(Linear(W_1,b_1,W_2,b_2,x),\alpha)=$$
$$=LeakyReLu(Linear(W_2,b_2,\tanh(Linear(W_1,b_1,x))),\alpha)$$

Where:

$$Linear(x)=W\cdot x+b$$

$$dw\,linear:\frac{\partial}{\partial W}[W\cdot x+b]=x$$

$$dx\,linear:\frac{\partial}{\partial x}[W\cdot x+b]=W$$

$$db\,linear:\frac{\partial}{\partial b}[W\cdot x+b]=1$$

$$\tanh(x)=\frac{e^x-e^{-x}}{e^x+e^{-x}}$$

$$\frac{\tanh(x)}{\delta x}=\delta x\,\frac{e^x-e^{-x}}{e^x+e^{-x}}=\frac{(e^x-e^{-x})\cdot(e^x+e^{-x})^{-1}}{\delta x}=\frac{(e^x-e^{-x})}{\delta x}\cdot(e^x+e^{-x})^{-1}+(e^x-e^{-x})\cdot\frac{(e^x+e^{-x})^{-1}}{\delta x}=$$
$$=\frac{(e^x-e^{-x})}{\partial(e^x-e^{-x})}\cdot\frac{(e^x-e^{-x})}{\partial x}\cdot(e^x+e^{-x})^{-1}+(e^x-e^{-x})\cdot\frac{(e^x+e^{-x})^{-1}}{\partial(e^x+e^{-x})}\cdot\frac{(e^x+e^{-x})^{-1}}{\partial x}=$$
$$=(e^x+e^{-x})\cdot(e^x+e^{-x})^{-1}-(e^x-e^{-x})\cdot(e^x+e^{-x})^{-2}\cdot(e^x-e^{-x})=1-(e^x-e^{-x})^2\cdot(e^x+e^{-x})^{-2}$$

$LeakyReLU(x)=\begin{cases} x, x>0 \\ \alpha\cdot x, x\leq 0\end{cases}$ here α is a slope, not the learning rate

$$\frac{LeakyReLU(x)}{\partial x}=\begin{cases} 1, x>0 \\ \alpha, x\leq 0\end{cases}$$

MAE loss function:

$$MAE(y', y)=\frac{1}{N}\sum_{i=1}^{N}(y_i-y_i')$$

SGD:

$$W_i'=W_i-\alpha\cdot\frac{MAE(y, W_1, b_1, W_2, b_2, x)}{\partial W_i}$$
$$b_i'=b_i-\alpha\cdot\frac{MAE(y, W_1, b_1, W_2, b_2, x)}{\partial b_i}$$

$$y'=M(x)=LeakyReLu(Linear(W_1, b_1, W_2, b_2, x), \alpha)=$$
$$=LeakyReLu(Linear(W_2, b_2, \tanh(Linear(W_1, b_1, x))), \alpha)$$

$$m=Linear(W_1, b_1, x)$$
$$k=Linear(W_2, b_2, \tanh(Linear(W_1, b_1, x)))$$
$$l=\tanh(Linear(W_1, b_1, x))$$

$$\frac{y'}{\partial W_2}=\frac{LeakyReLu(k)}{\partial k}\cdot\frac{k}{\partial W_2}=$$
$$=\frac{LeakyReLu(k)}{\partial k}\cdot\frac{Linear(W_2, b_2, l)}{\partial W_2}=$$
$$=\frac{LeakyReLu(k)}{\partial k}\cdot l=\frac{LeakyReLu(k)}{\partial k}\cdot\tanh(Linear(W_1, b_1, x))$$

$$\frac{y'}{\partial b_2}=\frac{LeakyReLu(k)}{\partial k}\cdot\frac{k}{\partial b_2}=\frac{LeakyReLu(k)}{\partial k}\cdot\frac{Linear(W_2, b_2, l)}{\partial b_2}=\frac{LeakyReLu(k)}{\partial k}\cdot 1$$

$$\frac{y'}{\partial W_1} = \frac{LeakyReLu(k)}{\partial k} \cdot \frac{k}{\partial W_1} =$$

$$= \frac{LeakyReLu(k)}{\partial k} \cdot \frac{Linear(W_2, b_2, l)}{\partial W_1} =$$

$$= \frac{LeakyReLu(k)}{\partial k} \cdot \frac{Linear(W_2, b_2, l)}{\partial l} \cdot \frac{l}{\partial W_1} =$$

$$= \frac{LeakyReLu(k)}{\partial k} \cdot \frac{Linear(W_2, b_2, l)}{\partial l} \cdot \frac{\tanh(Linear(W_1, b_1, x))}{\partial W_1} =$$

$$= \frac{LeakyReLu(k)}{\partial k} \cdot \frac{Linear(W_2, b_2, l)}{\partial l} \cdot \frac{\tanh(m)}{\partial m} \cdot \frac{m}{\partial W_1} =$$

$$= \frac{LeakyReLu(k)}{\partial k} \cdot \frac{Linear(W_2, b_2, l)}{\partial l} \cdot \frac{\tanh(m)}{\partial m} \cdot \frac{Linear(W_1, b_1, x)}{\partial W_1} =$$

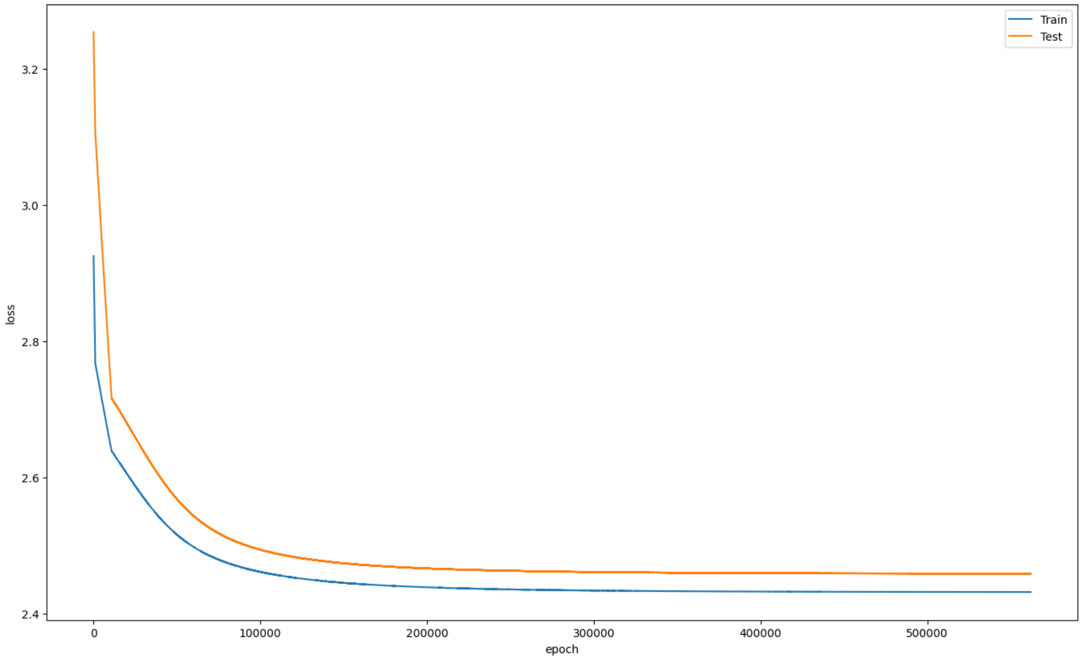
$$= \frac{LeakyReLu(k)}{\partial k} \cdot \frac{Linear(W_2, b_2, l)}{\partial l} \cdot \frac{\tanh(m)}{\partial m} \cdot x$$

$$\frac{y'}{\partial b_1} = \frac{LeakyReLu(k)}{\partial k} \cdot \frac{k}{\partial b_1} =$$

$$\ldots same\ as\ for\ W_1 \ldots$$

$$= \frac{LeakyReLu(k)}{\partial k} \cdot \frac{Linear(W_2, b_2, l)}{\partial l} \cdot \frac{\tanh(m)}{\partial m} \cdot 1$$

Running this model produces following result:


Train
Test
loss
epoch