1) $MSE = \frac{1}{N}(y-\hat{y})^2 \Rightarrow \frac{1}{N}\left(y - relu\left(\sum_{j=0}^{M} w_j x_j\right)\right)^2$

when $x<0$, $relu\left(\sum_{j=0}^{M} w_j x_j\right) = 0$

$MSE = \frac{1}{N}(y-0)^2 = \frac{1}{N}(y^2)$, this is a constant.

Thus when $x<0$ relu is not quadratic.

so $\frac{\partial f}{\partial w} = \frac{1}{N} y^2 = 0$

Gradient descent

$w \longleftarrow w - \alpha(0)$

$w \longleftarrow w$

The gradient does not change.

when $x>0$, $MSE = \frac{1}{N}\left(y - relu\left(\sum_{j=0}^{M} w_j x_j\right)\right)^2$

Let $u = relu\left(\sum_{j=0}^{M} w_j x_j\right)$

$MSE = \frac{1}{N}(y-u)^2 = \frac{1}{N}(y^2 - 2yu + u^2)$

$= \frac{1}{N}\left(y^T y - 2y\, relu\left(\sum_{j=0}^{M} w_j x_j\right) + relu\left(\sum_{j=0}^{M} w_j x_j\right)^2\right)$

As we can see this is quadratic.

$\frac{\partial f}{\partial w} = \frac{1}{N}\left(-2y\, relu \sum_{j=0}^{M} x_j + 2\, relu \sum_{j=0}^{M} w_j x_j\right)$

$= \frac{2}{N}\left(-y\, relu \sum_{j=0}^{M} x_j + relu \sum_{j=0}^{M} w_j \cdot relu \sum_{j=0}^{M} x_j\right)$

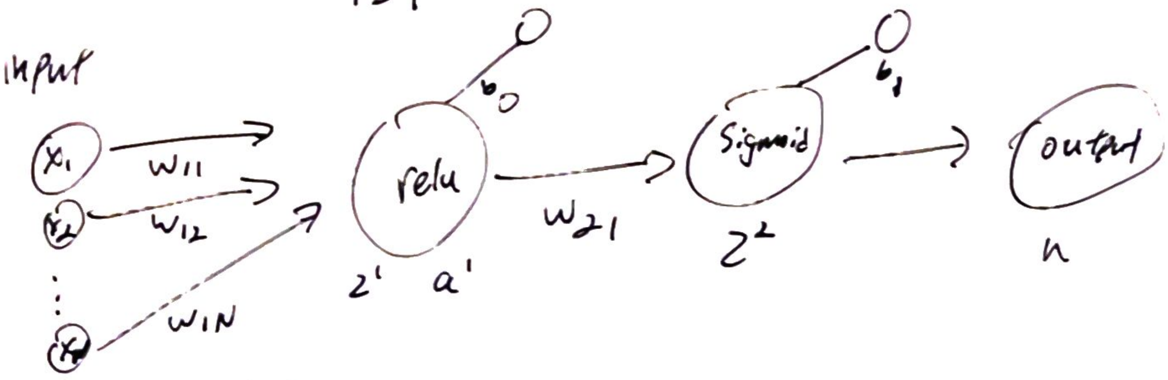
$= \frac{2}{N} relu \sum_{j=0}^{M} x_j \left(-y + relu \sum_{j=0}^{M} w_j\right) = 0$

$w^{*} = \left(\frac{2}{N} relu \sum_{j=0}^{M} x_j\right)^{-1} + y = relu \sum_{j=0}^{M} w_j$

Gradient descent

$w \longleftarrow w - \alpha\left(\frac{1}{N}\left(-2y\, relu \sum_{j=0}^{M} x_j + 2\, relu \sum_{j=0}^{M} w_j x_j\right)\right)$

2) $\text{LOSS} = \sum_{i=1}^{N} y_i \log(nn(x_i; w)) + (1-y_i) \log(1-nn(x_i; w))$



$$z_1 = \left(\sum_{i=1}^{N} w_{1N} \cdot x_N\right) + b_0$$

$$a_1 = \text{relu}(z_1)$$

$$\begin{aligned} z_2 &= w_{21} \cdot a^2 + b_1 \\ &= w_{21} \cdot \text{relu}(z_1) + b_0 \\ &= w_{21} \cdot \text{relu}\left(\left(\sum_{i=1}^{N} w_{1N} \cdot x_N\right) + b_0\right) + b_1 \end{aligned}$$

$$h = a_3 = \text{Sigmoid}(z_2) = \text{output} = nn(x_i; w_i)$$

$$J = \text{Loss}$$

$$\frac{\partial J}{\partial w} = \frac{\partial J}{\partial h} \cdot \frac{\partial h}{\partial z_2} \cdot \frac{\partial z_2}{\partial w}$$

Let $u = nn(x_i; w_i)$

$$\frac{\partial J}{\partial h} = \sum_{i=1}^{N} \frac{1}{u \ln} \cdot y_i + (1-y_i) \frac{1}{(1-u)\ln}$$

$$\frac{\partial h}{\partial z_2} = \sigma(z_2)(1-\sigma(z_2))$$

derivation of sigmoid

$$\frac{ds(x)}{dx} = \frac{1}{1+e^{-x}} = \left(\frac{1}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right)(-e^{-x}) = \sigma(x)(1-\sigma(x))$$

$$\frac{\partial z_2}{\partial w} = w_{21} \cdot relu\left(\sum_{i=1}^{N} w_{1N} \cdot x_N\right)$$

$$\frac{\partial J}{\partial w} = \sum_{i=1}^{N} \frac{1}{u \ln} \cdot y_i + (1-y_i)\frac{1}{(1-u)\ln} \cdot \sigma(z_2)(1-\sigma(z_2)) \cdot w_{21} relu\left(\sum_{i=1}^{N} w_{1N} \cdot x_N\right)$$

$$w \longleftarrow w - \alpha\left(\frac{\partial J}{\partial w}\right)$$

3) From Assignment 1
we know $w^T Q w + b w + y_i^T y_i = MSE$

so $w^T \frac{x^T x}{N} w - \frac{2y}{N} x w + \frac{y^T y}{N} + \lambda w^T w$

$$= w^T \frac{x^T x}{N} w + \lambda w^T w - \frac{2y}{N} x w + \frac{y^T y}{N}$$

$$= w^T \left(\frac{x^T x}{N} + \lambda I\right) w - \frac{2y}{N} x w + \frac{y^T y}{N}$$

I is an identity matrix and the loss is a Quadratic form

$$\nabla Loss = 2\left(\frac{x^T x}{N} + \lambda I\right) w + \frac{-2y}{N}(x) = 0$$

$$w = \frac{2y}{N}(x) \cdot \frac{1}{2}\left(\frac{x^T x}{N} + \lambda I\right)^{-1}$$