Exercise 1:

Sigmoid Function:

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\frac{\partial(\sigma(x))}{\partial(x)} = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}}(1 - \frac{1}{1+e^{-x}}) = \sigma(x)(1-\sigma(x)) \text{ } \mathbf{(1)}$$

Let:

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} \\ ... & ... & ... \\ 1 & x_1^{(n)} & x_2^{(n)} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}, w = \begin{bmatrix} w_o \\ w_1 \\ w2 \end{bmatrix}$$

We have Loss function:

$L = -\sum_{i=1}^{N}(y_i log(\hat{y_i}) + (1-y_i) log(1 \hat{-} y_i))$

Predict Function in Logistic Function:

$\hat{y_i} = \sigma(w_o + w_1 x_1^{(i)} + w_2 x_2^{(i)}) = \sigma(X.w)$

+ Gradient descent

Our purpose is minimize the Loss Function to find W.

From the chain Rule we have:

$$\frac{\partial(L)}{\partial(w)} = \left(\frac{\partial(\hat{y_i})}{\partial(w)}\right)^T . \frac{\partial(L)}{\partial(\hat{y_i})}$$

While:

$$\frac{\partial(L)}{\partial(\hat{y_i})} = -\left(\frac{y_i}{\hat{y_i}} - \frac{1-y_i}{1-\hat{y_i}}\right)$$

$$\frac{\partial(\hat{y_i})}{\partial(w_o)} = \frac{\partial(\sigma(w_o + w_1 x_1^{(i)} + w_2 x_2^{(i)}))}{\partial(w_o)} = \hat{y_i}(1-\hat{y_i})\mathbf{(1)}$$

$$\Rightarrow \frac{\partial(L)}{\partial(w_o)} = \frac{\partial(L)}{\partial(\hat{y_i})} . \frac{\partial(\hat{y_i})}{\partial(w_o)} = -\left(\frac{y_i}{\hat{y_i}} - \frac{1-y_i}{1-\hat{y_i}}\right) . \hat{y_i}(1-\hat{y_i})$$

$$= -\left[y_i(1-\hat{y_i}) - (1-y_i)\hat{y_i}\right] = (\hat{y_i} - y_i)$$

Similar with w_1 and w_2 we have:

$$\rightarrow \frac{\partial(L)}{\partial(w_1)} = \frac{\partial(L)}{\partial(\hat{y_i})} . \frac{\partial(\hat{y_i})}{\partial(w_1)} = -\left(\frac{y_i}{\hat{y_i}} - \frac{1-y_i}{1-\hat{y_i}}\right) . x_1^{(i)} \hat{y_i}(1-\hat{y_i}) = x_1^{(i)}(\hat{y_i} - y_i)$$

$$\rightarrow \frac{\partial(L)}{\partial(w_2)} = \frac{\partial(L)}{\partial(\hat{y}_i)}.\frac{\partial(\hat{y}_i)}{\partial(w_2)} = -\left(\frac{y_i}{\hat{y}_i} - \frac{1-y_i}{1-\hat{y}_i}\right).x_2^{(i)}\hat{y}_i(1-\hat{y}_i) = x_2^{(i)}(\hat{y}_i - y_i)$$

$$\Rightarrow \frac{\partial(L)}{\partial(w)} = \left(\frac{\partial(\hat{y}_i)}{\partial(w)}\right)^T.\frac{\partial(L)}{\partial(\hat{y}_i)} = \begin{bmatrix} \frac{\partial(L)}{\partial(w_o)} \\ \frac{\partial(L)}{\partial(w_1)} \\ \frac{\partial(L)}{\partial(w_2)} \end{bmatrix} = \begin{bmatrix} (\hat{y}_i - y_i) \\ x_1^{(i)}(\hat{y}_i - y_i) \\ x_2^{(i)}(\hat{y}_i - y_i) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & ... & 1 \\ x_1^1 & ... & x_1^n \\ x_2^1 & ... & x_2^n \end{bmatrix}.(\hat{y}_i - y_i) = X^T(\hat{y}_i - y_i)$$