LOGISTIC REGRESSION

Exercise 1. *Transform the logistic regression algorithm, from formula building, likelihood, maximize likelihood, negative log derivative according to coefficient matrix.*
Consider first of all the case of 2 classes. The ***posterior probability*** for class *C1* can be written as:

$$p(C1|x) = \frac{p(x|C1)p(C1)}{p(x|C1)p(C1) + p(x|C2)p(C2)} \tag{1}$$

Divide the numer and denominator by $p(x|C1)p(C1)$:

$$p(C1|x) = \frac{1}{1 + \frac{p(x|C2)p(C2)}{p(x|C1)p(C1)}} = \frac{1}{1 + e^{-a}} \tag{2}$$

$$where a = log \frac{p(x|C1)p(C1)}{p(x|C2)p(C2)} \tag{3}$$

So, the logistic sigmoid function is:

$$\sigma(a) = \frac{1}{1 + e^{-a}} \tag{4}$$

1 Calculate the derivatives of sigmoid function

$$\frac{d\sigma(x)}{dx} = \frac{d(\frac{1}{1+e^{-x}})}{dx} = \frac{-1}{(1 + e^{-x})^2}.1.e^{-x}.(-1) = \frac{e^{-x}}{(1 + e^{-x})^2} \tag{5}$$

$$= \frac{1}{1 + e^{-x}}.\frac{e^{-x}}{1 + e^{-x}} = \frac{1}{1 + e^{-x}}(1 - \frac{1}{1 + e^{-x}}) = \sigma(x)(1 - \sigma(x)) \tag{6}$$

We using sigmoid function to model probability. The model is defined as:
$p\left(C1|\varphi\right) = y\left(\varphi\right) = \sigma\left(\mathbf{w}^T\varphi\right)$ $(\varphi is transformation of input x)$

$$p\left(C2|\varphi\right) = 1 - p\left(C1|\varphi\right) \tag{7}$$

2 Likelihood function

For a data set φ_n, t_n, where $t_n \in \{0, 1\}$ and $\varphi_n = \varphi(x_n)$ with n = *1,...,N*
The likelihood function (based on Bernoulli distribution) is:

$$p\left(t|\mathbf{w}\right)=\prod_{n=1}^{N}y_n^{t_n}\left(1-y_n\right)^{1-t_n} \tag{8}$$

where $t=\left(t_n,...t_N\right)^T$ and $y_n=p\left(C1|\varphi_n\right)$. Concretely, $\varphi_i, t_i \to \sigma\left(\mathbf{w}^T\varphi\right)$ with
φ_i: ith data point, t_i is the actual label of data (0 or 1) and y_i is the probability that model predicted.

3 Maximize likelihood

We can define an error function by taking the negative logarithm of the likelihood. First, we consider *likehood of 1 data point*:

$$p\left(t_i|\mathbf{w}\right)=y_i^{t_i}\left(1-y_i\right)^{1-t_i} \tag{9}$$

Then we have *likelihood of data set* (all of data points):

$$p\left(t|\mathbf{w}\right)=\prod_{i=1}^{N}y_i^{t_i}\left(1-y_i\right)^{1-t_i} \tag{10}$$

We need to ***maximize*** $\mathbf{p}\left(\mathbf{t}|\mathbf{w}\right)$ to ***find w***.
Loss function has the form (binary cross-entropy):

$$L=-\log p\left(t|\mathbf{w}\right)=-\textstyle\sum_{i=1}^{N}\left(t_i\log y_i+\left(1-t_i\right)\log(1-y_i)\right)$$

with $y_i=\sigma\left(\mathbf{w}^T\varphi\right)$, where $\mathbf{w}^T\varphi=w_0+w_1\varphi_1+...+w_D\varphi_D$

4 Calculate the derivatives of Loss function with respect to w

Consider Loss function of *1 data point*:

$$L=-t\log y+\left(1-t\right)\log(1-y) \text{ where } y=\sigma(w_0+w_1\varphi_1+...+w_D\varphi_D)$$

(t is actual value (0 or 1) and y is the value (probability) that model predicts)
Let $z=w_0+w_1\varphi_1+...+w_D\varphi_D$. Therefore, $y=\sigma(z)$
Apply chain rule:

$$\frac{\partial L}{\partial w_0}=\frac{\partial L}{\partial y}.\frac{\partial y}{\partial z}.\frac{\partial z}{\partial w_0} \tag{11}$$

$$\frac{\partial L}{\partial y} = -\frac{\partial(t\log y + (1-t)\log(1-y))}{\partial y} = -(\frac{t}{y} - \frac{1-t}{1-y}) = \frac{y-t}{y(1-y)} \tag{12}$$

$$\frac{\partial y}{\partial z} = \sigma(z)(1-\sigma(z)) = y(1-y) \tag{13}$$

$$\frac{\partial z}{\partial w_0} = 1 \tag{14}$$

Therefore,

$$\frac{\partial L}{\partial w_0} = \frac{y-t}{y(1-y)}.y(1-y).1 = y-t \tag{15}$$

Similarly,

$$\frac{\partial L}{\partial w_1} = \frac{y-t}{y(1-y)}.y(1-y).\varphi_1 = (y-t)\varphi_1 \tag{16}$$

$$\frac{\partial L}{\partial w_2} = (y-t)\varphi_2 \tag{17}$$

, etc.

Then we have

$\frac{\partial L}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial L}{\partial w_0} \\ \frac{\partial L}{\partial w_1} \\ ... \\ \frac{\partial L}{\partial w_D} \end{bmatrix} = \begin{bmatrix} (y-t).1 \\ (y-t)\varphi_1 \\ ... \\ (y-t)\varphi_D \end{bmatrix} = (y-t)\varphi$ **,** $(y-t)$ is constant

Taking the gradient of the error function with respect to $\boldsymbol{w}$, we obtain

$$\nabla L = \sum_{n=1}^{N} (y_n - t_n)\,\varphi_n \tag{18}$$

Exercise 2. *Find f(x), given that f'(x) = f(x)(1 - f(x))*

We have

$$f'(x) = \frac{df\,(x)}{dx} = f\,(x)\,(1 - f\,(x)) \tag{19}$$

$$\leftrightarrow dx = \frac{df\,(x)}{f\,(x)\,(1 - f\,(x))} \tag{20}$$

$$\leftrightarrow \int dx = \int \frac{df\left(x\right)}{f\left(x\right)\left(1-f\left(x\right)\right)} \tag{21}$$

$$\leftrightarrow \int dx = \int (\frac{1}{f\left(x\right)} + \frac{1}{1-f\left(x\right)})df\left(x\right) \tag{22}$$

$$\leftrightarrow \int dx = \int \frac{1}{f\left(x\right)}df\left(x\right) + \int \frac{1}{1-f\left(x\right)}df\left(x\right) \tag{23}$$

$\leftrightarrow x = \log f\left(x\right) - \log(1-f\left(x\right))$

$$\leftrightarrow x = \log \frac{f\left(x\right)}{1-f\left(x\right)} \tag{24}$$

$$\leftrightarrow \frac{f\left(x\right)}{1-f\left(x\right)} = e^{x} \tag{25}$$

$$\leftrightarrow f\left(x\right) = e^{x} - e^{x}f\left(x\right) \tag{26}$$

$$\leftrightarrow 1 = \frac{e^{x}}{f\left(x\right)} - e^{x} \tag{27}$$

$$\leftrightarrow f\left(x\right) = \frac{e^{x}}{1+e^{x}} = \frac{1}{1+e^{-x}} \tag{28}$$