1 Logistic Regressions:

The goal in classification is to take an input vector x and to assign it to one of K discrete classes C_K where k = 1,...,K.

The input space is thereby divided into decision regions whose boundaries are called decision boundaries or decision surfaces.

Consider the first of all the case of two classes. The posterior probability for class C_1 can be written as:

$$p(C_1|x) = \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1)+p(x|C_2)p(C_2)} = \frac{1}{1+e^{-a}} = \sigma(a)$$

where we defined:

$$a = \log \frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)}$$

Thus, we have the sigmoid function defined as:

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

Applying the quotient rule, we have the derivative of the sigmoid function as:

$$\begin{aligned}
\Rightarrow \frac{d}{dx}\sigma(x) &= \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right] \\
&= \frac{(0)(1+e^{-x}) - (-e^{-x})(1)}{(1+e^{-x})^2} \\
&= \frac{e^{-x}}{(1+e^{-x})^2} \\
&= \frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}}\frac{e^{-x}+(1-1)}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}}\frac{(1+e^{-x})-1}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}}\left[\frac{(1+e^{-x})}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right] \\
&= \frac{1}{1+e^{-x}}\left[1 - \frac{1}{1+e^{-x}}\right] \\
&= \sigma(x)(1-\sigma(x))
\end{aligned}$$

The model logistic regression is defined as:

$$p(C_1|\phi) = y(\phi) = \sigma(\mathbf{w^T}\phi)$$

$$p(C_2|\phi) = 1 - p(C_1|\phi)$$

For a data set ϕ_n, t_n, where $t_n \in \{0, 1\}$ and $\phi_n = \phi(X_n)$, with n = 1,...,N, thelikelihood function can be written as:

$$p(t_n|w) = y_n^{t_n}(1 - y_n)^{1-t_n}$$

$$p(t|w) = \sum_{n=1}^{N} y_n^{t_n}(1 - y_n)^{1-t_n}$$

where t = $(t_1, ...t_N)^\mathbf{T}$ and $y_n = p(C_1|\phi_n) = \sigma(\mathbf{w^T}\phi)$

We can define an error function by taking the negative logarithm of thelikelihood, which gives the cross-entropy error function in theform

$$L = -\log p(t|w) = -\sum_{n=1}^{N}\{t_n \log y_n + (1 - t_n)\log(1 - y_n)\}$$

where $y_n = \sigma(a_n)$ and $a = \mathbf{w^T}\phi_n$

Taking the gradient of the error function with respect to w, we obtain

Applying the chain rule with z = a_n where $a = \begin{bmatrix} W_0 \\ W_1\phi_1 \\ \cdot \\ \cdot \\ W_n\phi_n \end{bmatrix}$ and y = $\sigma(z)$, we have:

$$
\begin{aligned}
\frac{\partial L}{\partial w} &= \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z} \times \frac{\partial z}{\partial w} \\
&= (-\frac{t}{y} + \frac{(1-t)\times(-1)}{1-y}) \times \sigma(z)(1-\sigma(z)) \times \phi_1 \\
&= (\frac{1-t}{1-y} - \frac{t}{y}) \times \sigma(z)(1-\sigma(z)) \times \phi_1 \\
&= (\frac{y - ty - t + yt}{(1-y)y}) \times \sigma(z)(1-\sigma(z)) \times \phi_1 \\
&= \frac{y-t}{(1-y)y} \times \sigma(z)(1-\sigma(z)) \times \phi_1 \\
&= \frac{y-t}{(1-y)y} \times (1-y)y \times \phi_1 \\
&= (y-t) \times \phi_1 \\
&= \sum_{n=1}^{N} (y_n - t_n)\phi_n
\end{aligned}
$$

2 Find f(x), knowing f'(x) = f(x)(1-f(x))

From the results above of the derivative of sigmoid function:

$$\sigma'(a_n) = \sigma(x)(1-\sigma(x))$$

We can safely conclude that f(x) in this question can be definitively be in the form of : $\frac{1}{1+e^{-a}}$