For a given dataset $D = \{x_i, y_i\}_{i=1}^N$ with $x_i \in R^D$ and $y_i \in \{0, 1\}$, consider the following objective function $L(w)$:

With logistic function

$$L(w) = \sum_{i=1}^{N} \left[f(w^T x_i) - y_i \right]^2 \tag{1}$$

and

$$f(t) = \frac{1}{1 + \exp(-t)} \tag{2}$$

Show that $L(w^*)$ has a minimum value, not a maximum, at the w^* making

$$\left. \frac{\partial L}{\partial w} \right|_{w=w^*} = 0. \tag{3}$$

Let's go through the full process of calculating the **second derivative** of the objective function $L(w)$. This involves deriving the **Hessian matrix** of $L(w)$, which will help us determine whether $L(w)$ is convex and has a minimum at $w = w^*$. The objective function is given by:

$$L(w) = \sum_{i=1}^{N} \left[f(w^T x_i) - y_i \right]^2$$

where the logistic function $f(t)$ is:

$$f(t) = \frac{1}{1 + \exp(-t)}$$

We already derived the first derivative of $L(w)$:

$$\frac{\partial L}{\partial w} = 2 \sum_{i=1}^{N} \left[f(w^T x_i) - y_i \right] \cdot f(w^T x_i)(1 - f(w^T x_i)) x_i$$

Our goal now is to differentiate this expression with respect to w again to get the **second derivative**, or Hessian matrix H. We need to differentiate each term in the gradient with respect to w. Since the gradient involves a sum over i, the second derivative will also be a sum over i. The gradient has the form:

$$\frac{\partial L}{\partial w} = 2 \sum_{i=1}^{N} \left[g(w^T x_i) \right] x_i$$

where:

$$g(w^T x_i) = \left(f(w^T x_i) - y_i \right) f(w^T x_i)(1 - f(w^T x_i))$$

We need to compute the derivative of $g(w^T x_i)x_i$ with respect to w. Let's go step by step. We first differentiate $g(w^T x_i)$ with respect to w. Recall that:

$$f(w^T x_i) = \frac{1}{1 + \exp(-w^T x_i)}$$

and:

$$\frac{d}{dt} f(t) = f(t)(1 - f(t))$$

Using the chain rule, the derivative of $f(w^T x_i)$ with respect to w is:

$$\frac{\partial}{\partial w} f(w^T x_i) = f(w^T x_i)(1 - f(w^T x_i))x_i$$

Now, let's differentiate $g(w^T x_i)$ with respect to w. Applying the product rule:

$$\frac{\partial}{\partial w} \left[g(w^T x_i) \right] = \frac{\partial}{\partial w} \left(\left[f(w^T x_i) - y_i \right] f(w^T x_i)(1 - f(w^T x_i)) \right)$$

We treat $\left[f(w^T x_i) - y_i \right]$ and $f(w^T x_i)(1 - f(w^T x_i))$ as products. Differentiating the product of these two terms with respect to w, we apply the product rule:

$$\frac{\partial}{\partial w} \left(\left[f(w^T x_i) - y_i \right] f(w^T x_i)(1 - f(w^T x_i)) \right)$$

results in two terms:

1. Differentiate $\left[f(w^T x_i) - y_i \right]$:

$$\frac{\partial}{\partial w} \left[f(w^T x_i) - y_i \right] = f(w^T x_i)(1 - f(w^T x_i))x_i$$

2. Differentiate $f(w^T x_i)(1 - f(w^T x_i))$ using the chain rule. This gives a more complex expression but ultimately remains proportional to x_i.

Finally, the **Hessian matrix** is:

$$H = \frac{\partial^2 L}{\partial w^2} = 2 \sum_{i=1}^{N} \left[\frac{\partial}{\partial w} g(w^T x_i) \cdot x_i^T \right]$$

This results in a matrix because of the outer product of the vectors x_i and x_i^T.

- The second derivative, or Hessian, involves terms of the form $f(w^T x_i)(1 - f(w^T x_i))$, which are non-negative.
- Since the logistic function's second derivative is always non-negative, the Hessian matrix is **positive semi-definite**. This indicates that the function is **convex**, and thus, the critical point w^* is a **minimum**.

Through this full differentiation process, we have confirmed that the objective function $L(w)$ is convex, as the Hessian matrix is positive semi-definite. Therefore, w^* is indeed a **minimum**.