1. Introducción

El conjunto de datos MNIST (Fig. 1), contiene números escritos a mano que usualmente se usan para entrenar algoritmos de procesamiento de imágenes. En este trabajo, se utilizó este conjunto de datos con 50,000 filas y 784 columnas, con el objetivo de utilizar dos métodos, el método de gradiente descendente y el método de Newton modificado para maximizar la función.

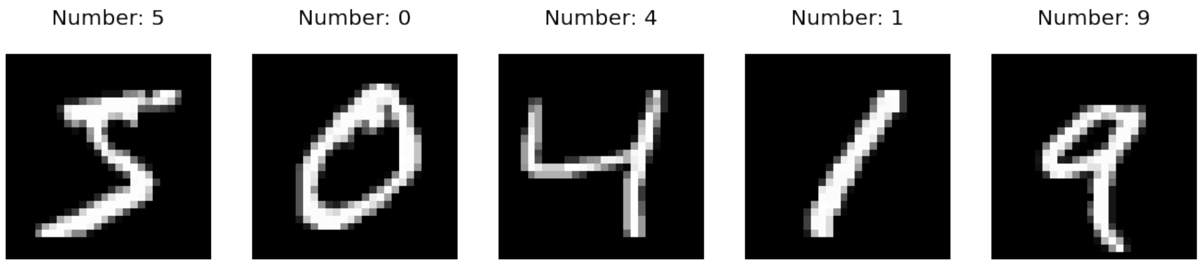

Figura 1: Imágenes del conjunto de datos MNIST.

Se utilizó la función logística de regresión con todos los signos cambiados,

$$h(\boldsymbol{\beta}, \beta_0) = \sum_{i=1}^{n} y_i log(\pi_i) + (1 - y_i) log(1 - \pi_i),$$

donde π_i

$$\pi_i(\boldsymbol{\beta}, \beta_0) = \frac{1}{1 + exp(-\mathbf{x}_i^T \boldsymbol{\beta} - \beta_0)}.$$

2. Metodología

Para calcular el método de gradiente descendente, se necesita la primera derivada de la función con respecto a β, y con respecto a β_0,

$$\frac{\partial h}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \mathbf{x}_i^T (y_i - \pi_i),$$
$$\frac{\partial h}{\partial \beta_0} = \sum_{i=1}^{n} (y_i - \pi_i),$$

por lo que el método de gradiente descendente sería:

Algorithm 1 Gradiente descendente

for iteraciones **do**
    z = $\mathbf{x} \cdot \boldsymbol{\beta} + \beta_0$
    $\pi = 1{,}0/(1 + exp(-z))$
    gradient = $\mathbf{x}^T \cdot (\mathbf{y} - \boldsymbol{\pi})$
    gradient_theta0 = suma($\mathbf{y} - \boldsymbol{\pi}$)
    $\theta - = \alpha$*gradient,
    $\theta_0 - = \alpha$*gradient_theta0

Para este problema se utiliza el algoritmo de Newton modificado porque al calcular el Hessiano de la función, se obtiene una matriz singular, por lo tanto no es invertible y no se podrá encontrar el tamaño de paso para la función. Por ello, se utilizó una matriz diagonal, y en todos los elementos se tenía la norma del Hessiano. Así, se evita la singularidad y se puede encontrar el tamaño de paso.

Algorithm 2 Newton Modificado

for iteraciones **do**
  Calcular z, y π como se calculó con el gradiente descendente,
  W = $\pi(1-\pi)$
  Hessiano = matmul((x.T, W), x)
  Calcular la norma del Hessiano y sumarla a los elementos de la diagonal
  gradient = $\mathbf{x}^T \cdot (\mathbf{y} - \boldsymbol{\pi})$
  gradient_theta0 = suma($\mathbf{y} - \boldsymbol{\pi}$)
  step = Hessiano^{-1} $\cdot$ gradient

  Ahora, calcular para theta0

3. Resultados

Se muestran los resultados obtenidos del error al comprobarlo con el conjunto de prueba, este se disminuyó porque al hacer las multiplicaciones de 50,000 por 50,000 la memoria era insuficiente, por lo que se utilizó un conjunto de (2000, 784).

	Iteraciones	Promedio error	Mínimo	Máximo
Gradiente descendente	100	3.6910	0.0398	8.0458
Newton Modificado	20	3.7297	3.6566	4.4496

Cuadro 1: Resultados de los dos métodos

4. Conclusiones

Se observó que la norma del gradiente no disminuía, aumentaba un poco en cada iteración o al utilizar un tamaño de paso para el gradiente descendente más grande, la norma crecía muy rápido por el término exponencial. Sin embargo, si había un error, no se encontró en el proceso. Asimismo, el método de Newton modificado, no se podía con las multiplicaciones más grandes, y por ello se decidió disminuir el tamaño. No obstante, con un buen tamaño de paso y también un conjunto de prueba aceptable, la norma se mantenía menor a uno.