

Considera la siguiente función de dos variables, que se asemeja a una función de pérdida cuadrática usada en redes neuronales:

$$f(w_1, w_2) = \frac{1}{2}(w_1x_1 + w_2x_2 - y)^2,$$

donde:

- x_1 y x_2 son valores de entrada,
- w_1 y w_2 son parámetros ajustables (pesos),
- y es el valor deseado (*target*).

Esta función es la pérdida cuadrática típica usada en el entrenamiento de una red neuronal simple (perceptrón sin activación).

Supón que:

- Entradas $\mathbf{x} = (x_1, x_2) = (1, 2)$,
- Etiqueta deseada $y = 4$.

Entonces, la función se reduce a:

$$f(w_1, w_2) = \frac{1}{2}(w_1 + 2w_2 - 4)^2.$$

1. Calcula el gradiente y los mínimos de la función $\nabla f(w_1, w_2)$: Deriva $f(w_1, w_2)$ respecto a w_1 y w_2 .
2. Analiza la matriz Hessiana de la función:
 - Calcula la matriz Hessiana H .
 - Encuentra sus valores propios y vectores propios.
3. Implementa el método de gradiente descendente para encontrar los valores óptimos de w_1 y w_2 que minimicen la función, calcula las primeras 5 iteraciones a mano:

$$w^{(k+1)} = w^{(k)} - t\nabla f(w^{(k)}),$$

y muestra los valores obtenidos en cada paso. Toma como punto inicial $w^{(0)} = (w_1^{(0)}, w_2^{(0)}) = (0, 0)$, y $t = 0,1$.

4. Calcula el gradiente y grafica el valor de f usando una escala logarítmica, para cada uno de las siguientes tasas de aprendizaje $lr = [0,005, 0,01, 0,05, 0,07, 0,1]$. Realiza 30 iteraciones por cada caso. Además, construye una tabla que muestre los valores finales de w_1 , w_2 y $f(w_1, w_2)$ correspondientes a cada caso.