

# Problema 5: Descenso de Gradiente y Descenso de Gradiente Basado en Momento

Marco Alejandro Ramirez - Juan Sebastian Sanchez

March 21, 2025

## 1. Introducción

El objetivo de este informe es comparar el rendimiento del método de descenso del gradiente y el descenso del gradiente con momento en la aproximación de la función seno. Se evaluará la velocidad de convergencia y la calidad de la aproximación obtenida.

## 2. Metodología

Para la aproximación de la función seno, se entrenó un modelo basado en regresión con descenso del gradiente estándar y con momento. La función de costo utilizada fue el error cuadrático medio (MSE), y la actualización de los parámetros se realizó según las siguientes ecuaciones:

- Descenso del Gradiente:

$$\theta_{t+1} = \theta_t - \alpha \nabla J(\theta_t) \quad (1)$$

- Descenso del Gradiente con Momento:

$$v_t = \beta v_{t-1} + (1 - \beta) \nabla J(\theta_t) \quad (2)$$

$$\theta_{t+1} = \theta_t - \alpha v_t \quad (3)$$

donde  $\alpha$  es la tasa de aprendizaje y  $\beta$  es el coeficiente de momento.

## 3. Descenso del Gradiente Estándar

El descenso del gradiente estándar actualiza los parámetros de la red neuronal siguiendo la dirección del gradiente:

1. **Inicialización:** Se crean acumuladores de gradientes en ceros.
2. **Cálculo del Gradiente:** Para cada muestra  $(x, y)$  en el mini-batch, se obtiene el gradiente mediante backpropagation.
3. **Promediado:** Se divide por el tamaño del mini-batch.
4. **Actualización de Parámetros:**

$$w = w - \eta \nabla J, \quad b = b - \eta \nabla J$$

donde  $\eta$  es la tasa de aprendizaje y  $\nabla J$  es el gradiente.

## 4. Descenso del Gradiente con Momentum

Este método introduce una velocidad  $v$  para mejorar la convergencia:

1. **Inicialización:** Se crean acumuladores de gradientes en ceros y velocidades iniciales en cero.
2. **Cálculo del Gradiente:** Se obtiene  $\nabla J$  como en el método estándar.
3. **Promediado:** Se divide por el tamaño del mini-batch.
4. **Actualización con Momentum:**

$$v_t = \mu v_{t-1} - \eta \nabla J$$
$$w = w + v_t, \quad b = b + v_t$$

donde  $\mu$  es el coeficiente de momentum (típicamente 0,9).

## 5. Resultados

Se generaron las siguientes gráficas para comparar la evolución del error y la aproximación obtenida antes y después del entrenamiento:

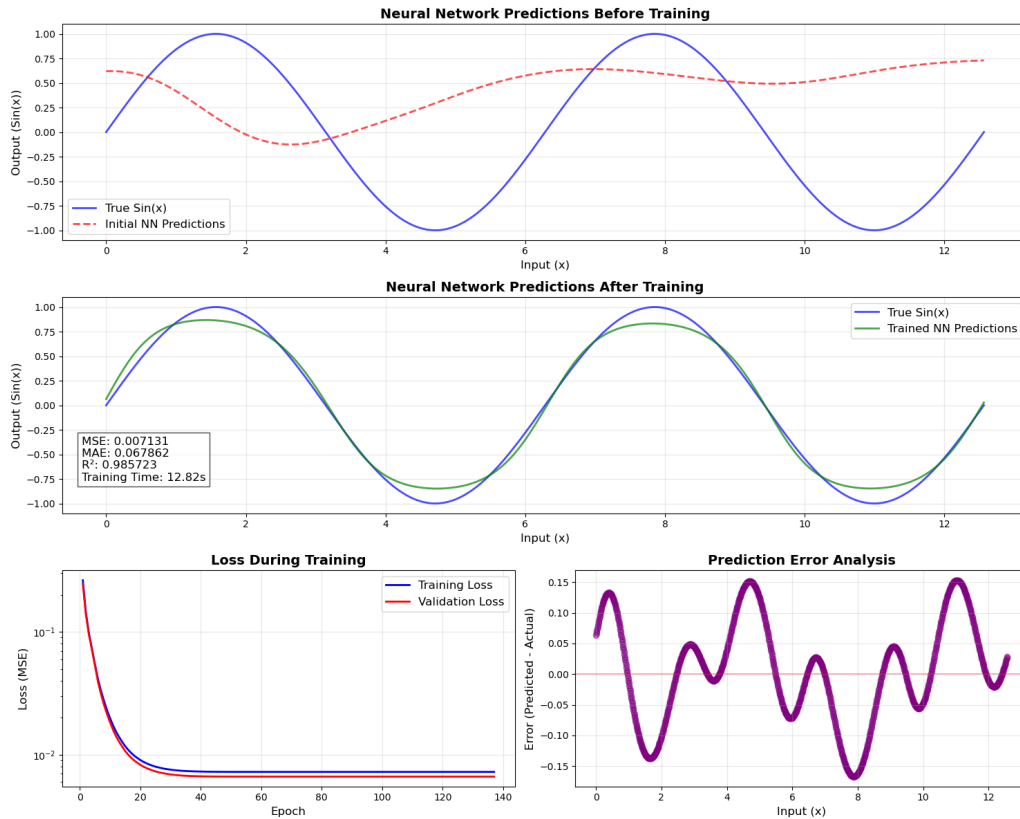


Figura 1: Predicciones, función de pérdida y predicción del error para Descenso del gradiente.

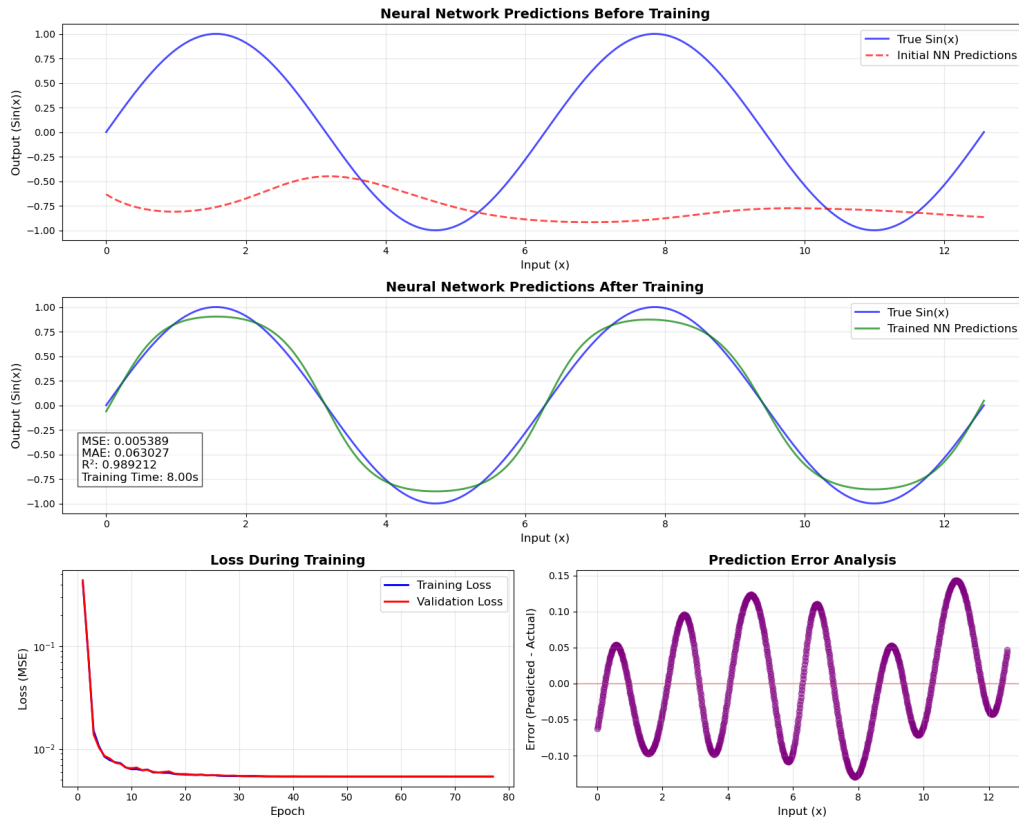


Figura 2: Predicciones, función de pérdida y predicción del error para Descenso del gradiente con momento.

## 6. Análisis Comparativo

### 6.1. Velocidad de convergencia

El descenso del gradiente con momento mostró una convergencia más rápida en comparación con el descenso del gradiente estándar. Esto se debe a que el término de momento ayuda a suavizar la trayectoria del descenso, evitando oscilaciones en direcciones con mayor curvatura. El Descenso de Gradiente Basado en Momento converge en aproximadamente 8 segundos y el clásico en 13 segundos.

### 6.2. Calidad de la aproximación

Ambos métodos lograron aproximar la función seno con buena precisión, pero el descenso del gradiente con momento alcanzó una mejor precisión en menos iteraciones. Además, la predicción del error también es superior para el método basado en momento. En las siguientes gráficas podemos ver más a detalle la predicción de la función seno para ambos métodos.

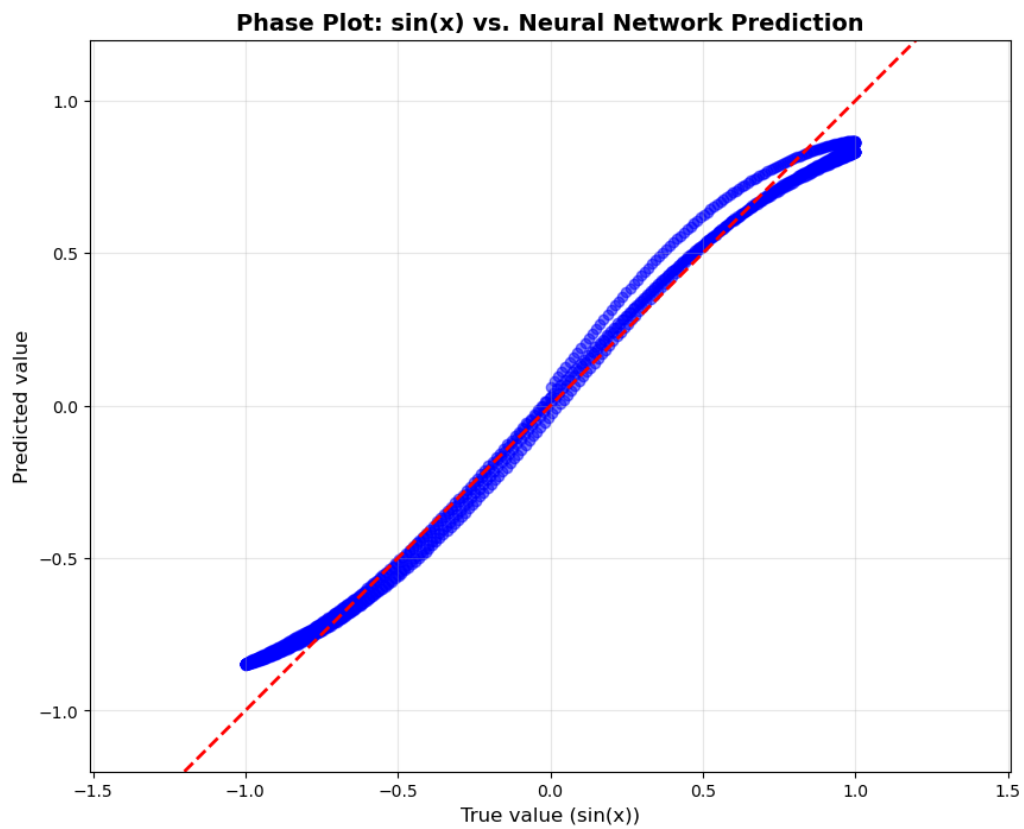


Figura 3: Plot de la función seno vs la predicción de la red neuronal entrenada con descenso del gradiente

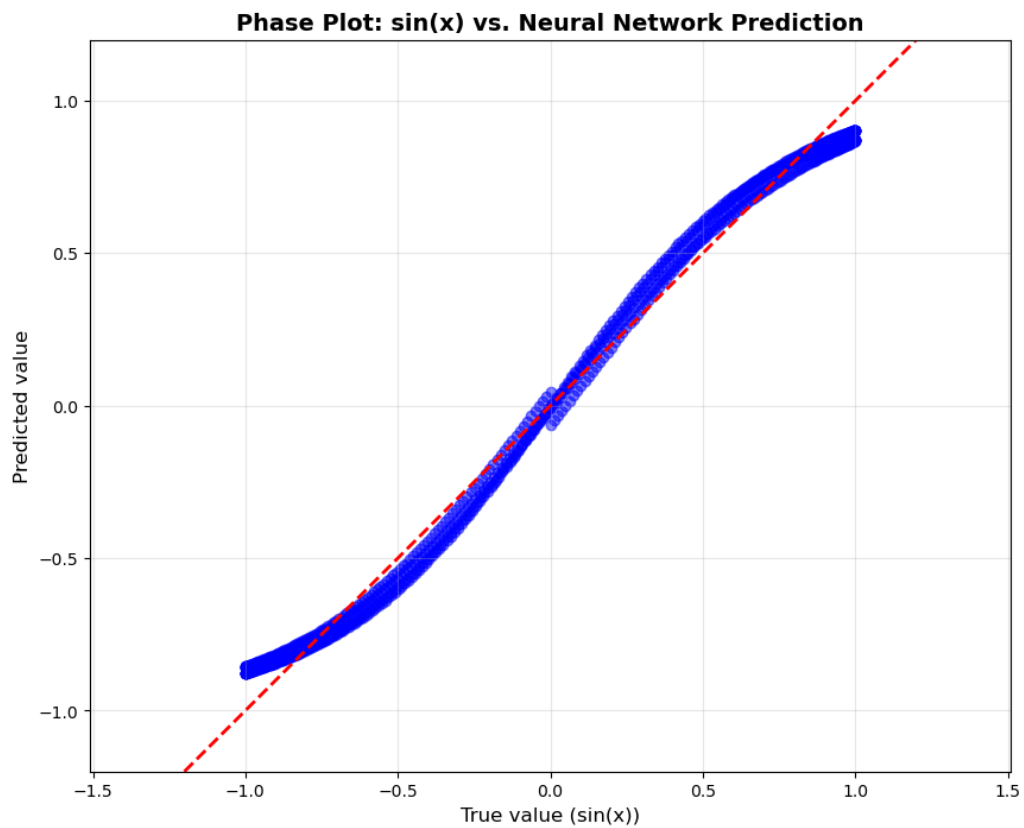


Figura 4: Plot de la función seno vs la predicción de la red neuronal entrenada con descenso del gradiente basado en momento

## 7. Conclusión

Se observó que el descenso del gradiente con momento mejora tanto la velocidad de convergencia como la calidad de la aproximación en comparación con el descenso del gradiente estándar. Esto lo convierte en una mejor opción para problemas de optimización con funciones no lineales.