

# Optimization Function

---

**Es solo el Método Matemático con el que se Optimizan los Parámetros de la ANN**

# Parameters of an ANN

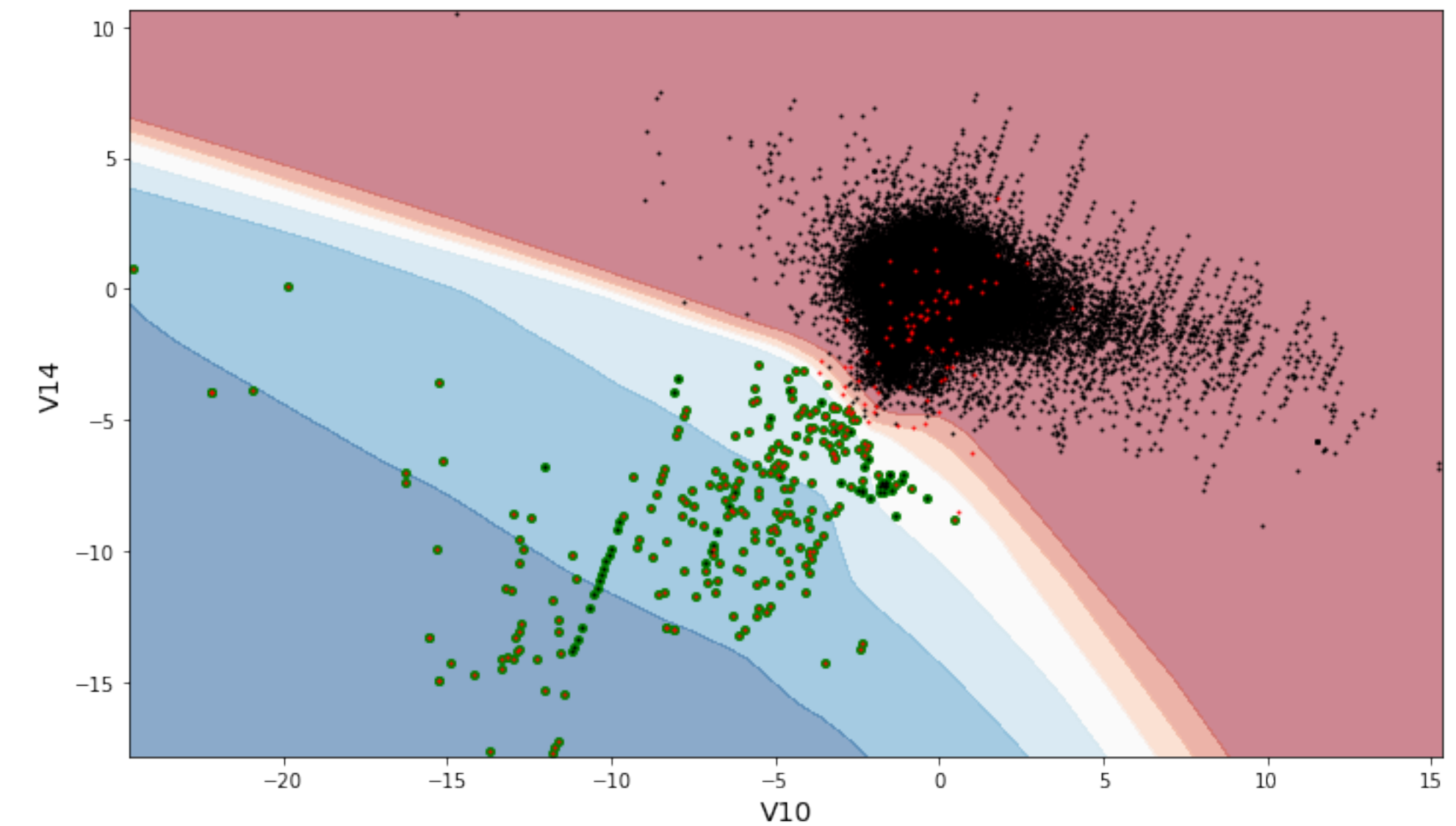
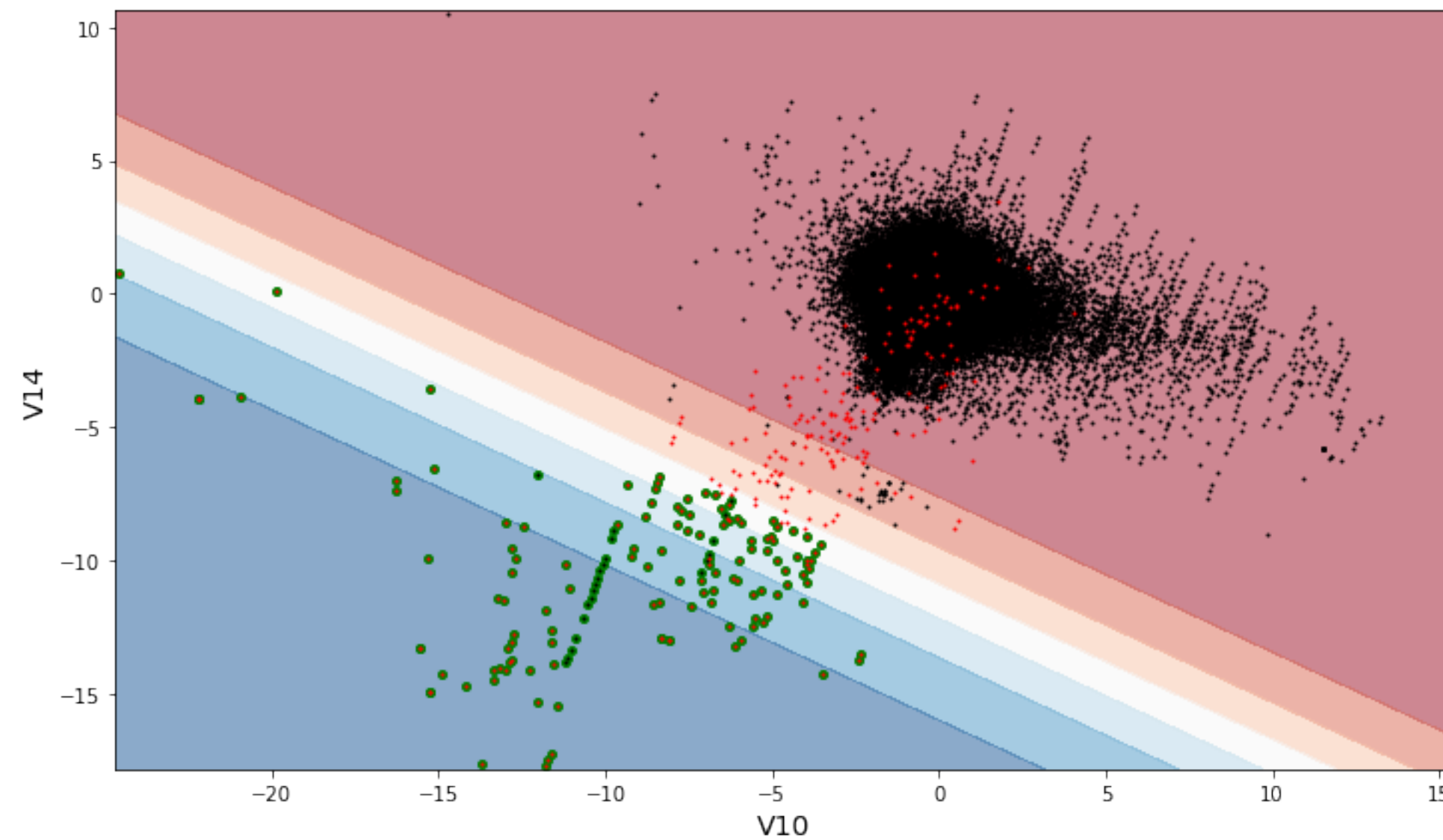
---

**En Conjunto son solo una Función Matemática**  
que clasifica o predice segmentando un **Decision**  
**Limit (Hypothesis Function)**

# Decision Limit (Hypothesis Function)

De acuerdo al **Número de Capas** (layers) definidas en el Modelo de ANN

A **Mayor** cantidad de **Capas (layer)** **Mayor Flexibilidad**



A **Mayor** cantidad de **Capas (layer)** **Mayor problemas de Overfitting**

# Batch Gradient Descent

---

Calculo de Derivadas de las Funciones y los Parámetros de la **ANN** por medio de un **Bucle Iterativo**

---

## Bucle Iterativo Unidimensional

Se computa todo el conjunto de datos para optimizar los parámetros de la **ANN**

---

El **bucle iterativo** se efectúa respecto a cada **Epoch** (computo total del DataSet)

Se vuelve **lento** la **Optimización** de la **ANN** con **DataSet** muy **grande** debido a la cantidad de **datos acumulados** en cada operación

# Mini-Batch Gradient Descent

---

**Calculo de Derivadas de las Funciones y los Parámetros de la ANN por medio de un Bucle Iterativo**

---

## **Bucle Iterativo Bidimensional**

Con cada Mini-Batch computado se optimizan los parámetros de la ANN

---

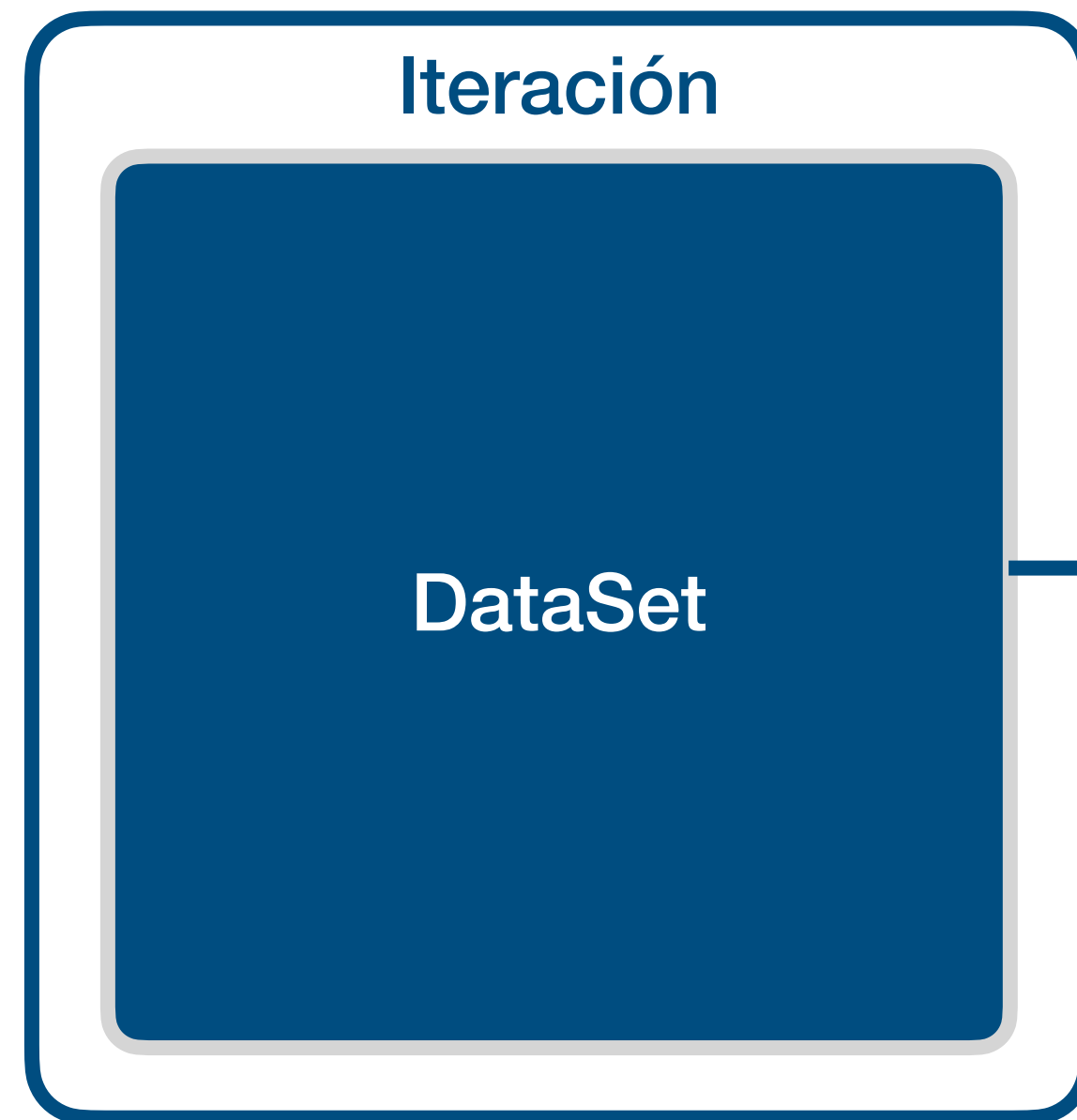
**El bucle iterativo se efectúa respecto a cada Mini-Batch que divide el conjunto de datos**

Al procesar todos los datos de un **Mini-Batch**, se determinan las **variables y estadísticos** y se **optimizan los parámetros de la ANN**

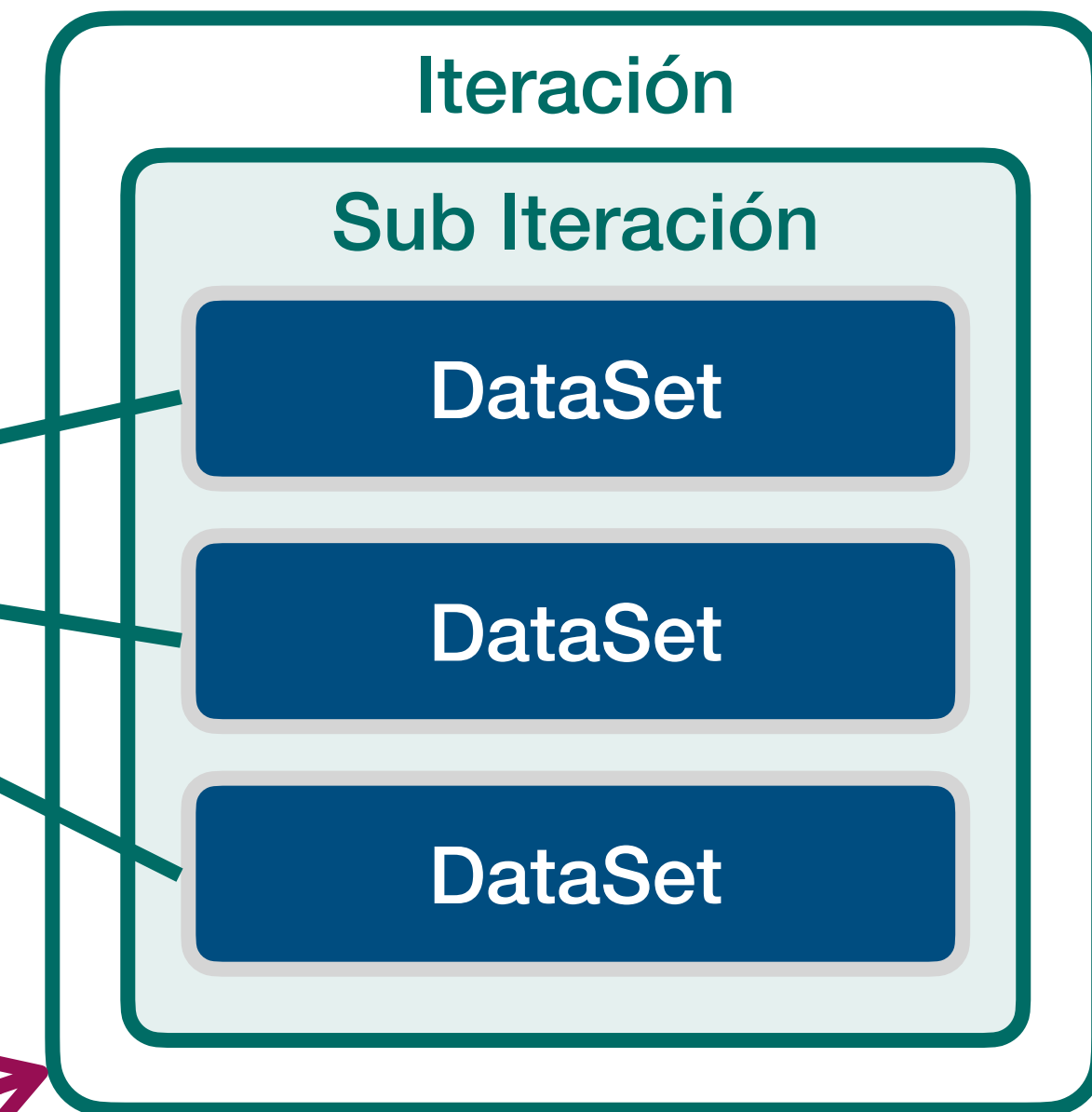
# Comparative

---

## Batch Gradient Descent



## Mini-Batch Gradient Descent



---

**Epoch** → **Recorrido completo del DataSet**

# Stochastic Gradient Descent

---

**Calculo de Derivadas de las Funciones y los Parámetros de la ANN por medio de un Bucle Iterativo**

---

## **Bucle Iterativo Bidimensional**

Con cada dato del conjunto de datos computado se optimizan los parámetros de la ANN

---

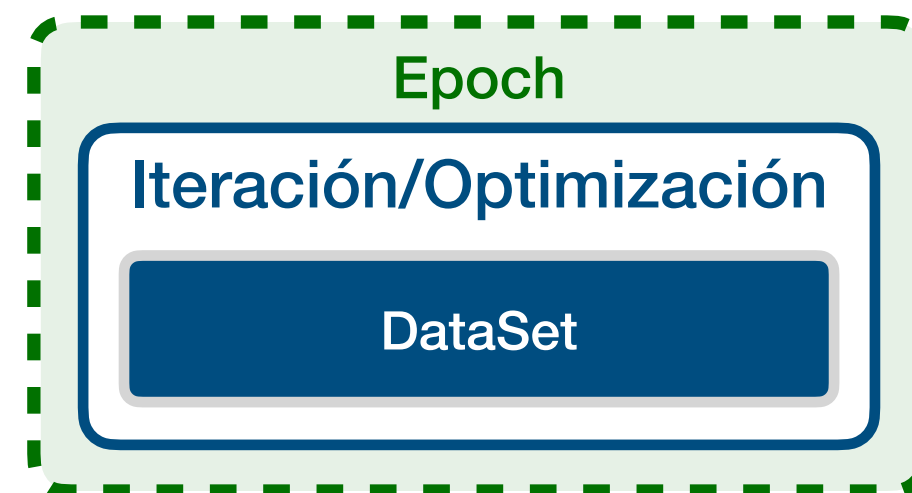
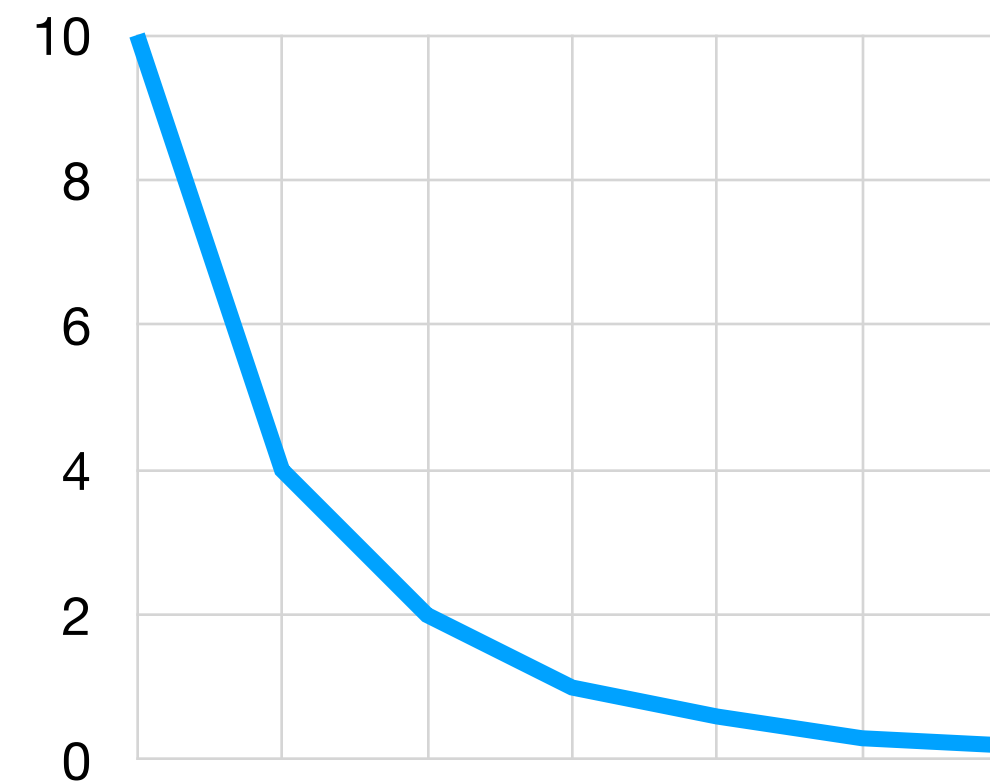
**El bucle iterativo se efectúa respecto a cada dato del conjunto de datos del DataSet**

**No suele converger con el valor de error mínimo de la Función de Error/Coste**



# Comparative

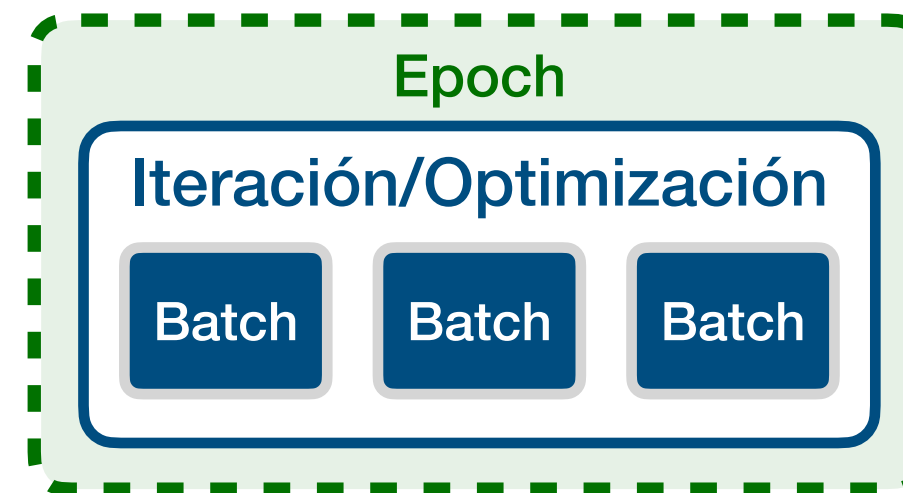
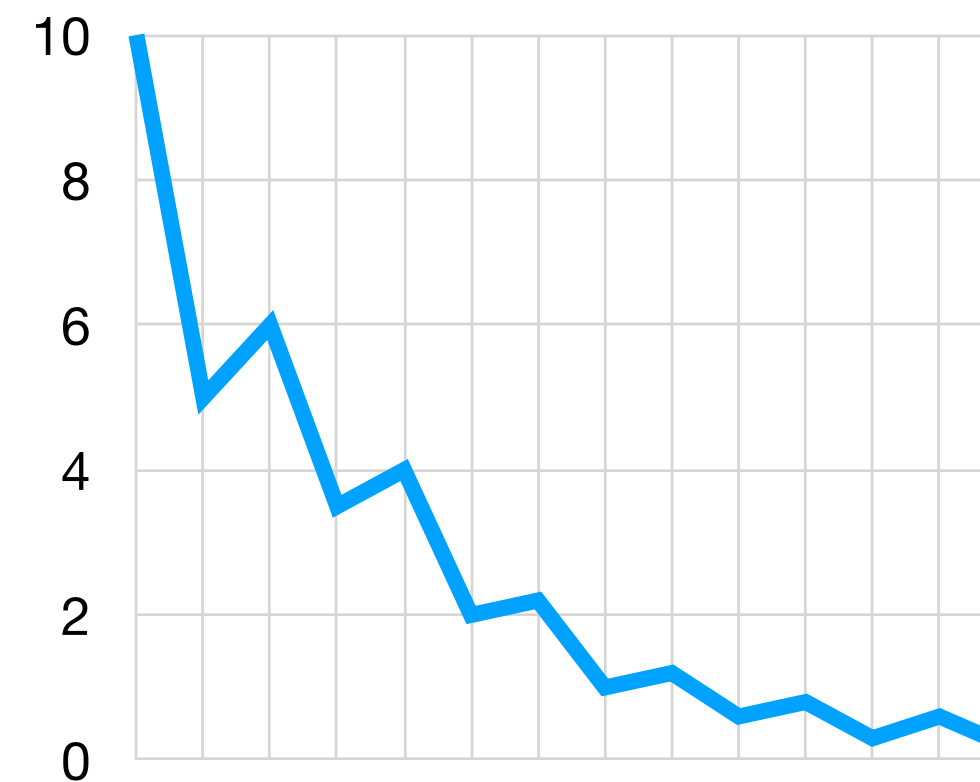
## Batch Gradient Descent



Por **cada conjunto de datos completo computado** se efectúa una **optimización** de los **parámetros** de la **ANN**

**Velocidad Lento**

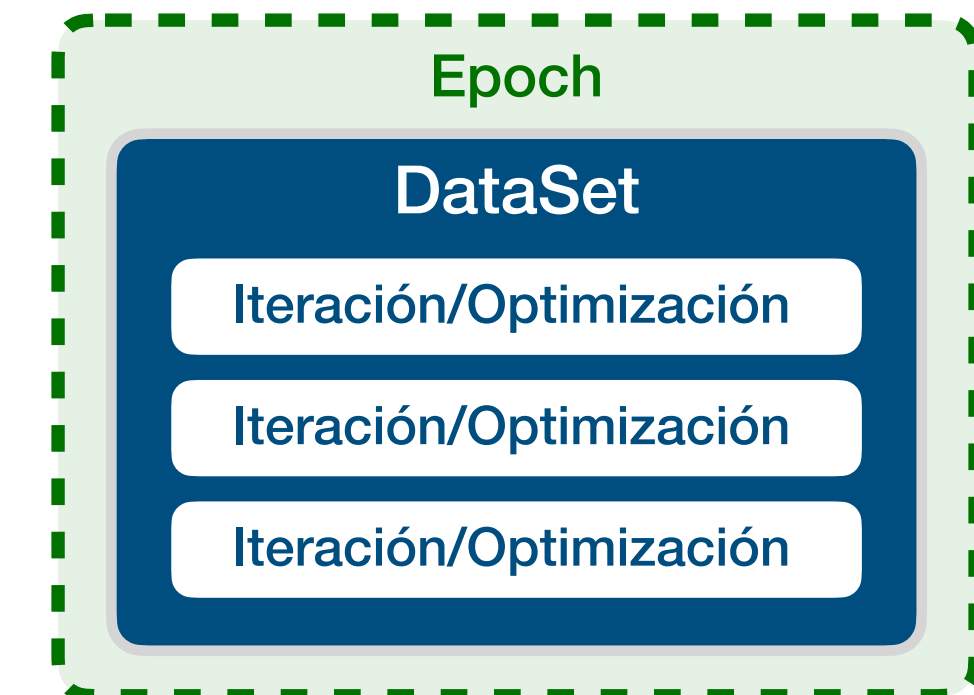
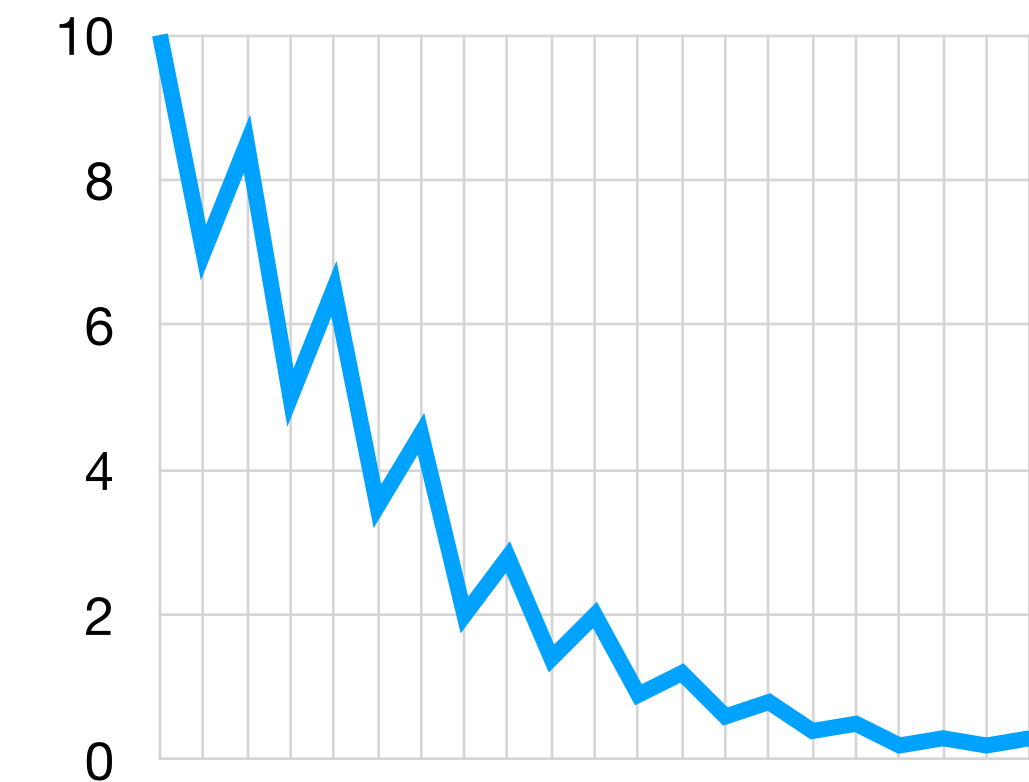
## Mini-Batch Gradient Descent



Por **cada Batch computado** del **conjunto de datos** se efectúa una **optimización** de los **parámetros** de la **ANN**

**Velocidad Media**

## Stochastic Gradient Descent



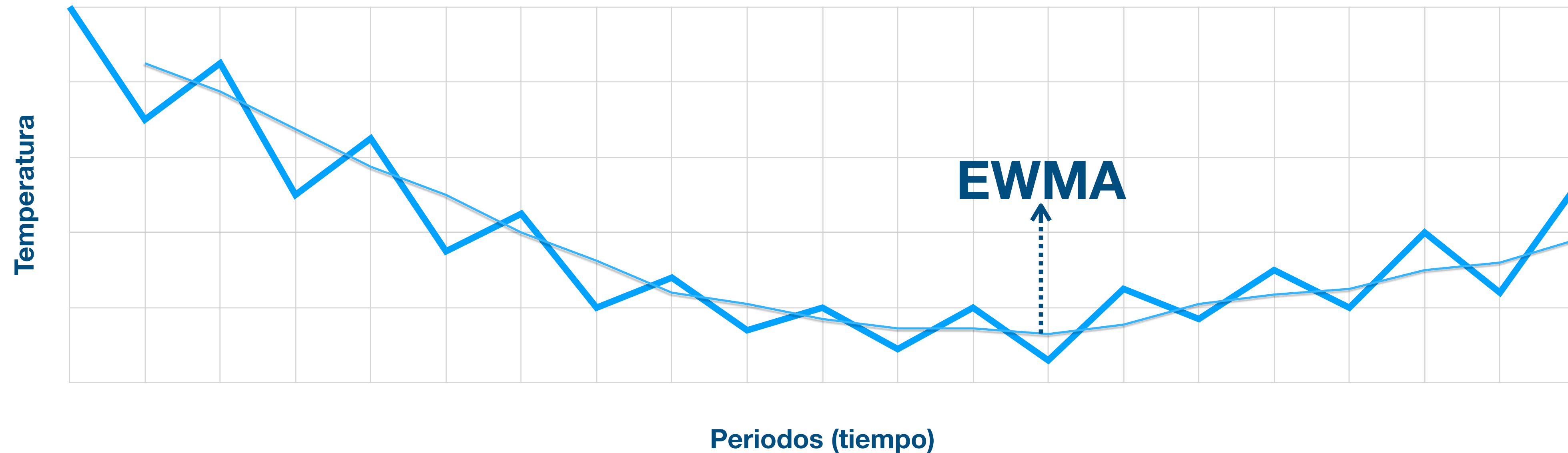
Por **cada dato computado** del **conjunto de datos** se efectúa una **optimización** de los **parámetros** de la **ANN**

**Velocidad Rapido**



# Exponentially Weighted Moving Average

---



---

**Calculo de la media de los datos pasados en conjunto  
para evaluar la media del nuevo dato a predecir**

# Gradient Descent with Momentum

---



## Algoritmo de Optimización para el algoritmo Mini-Batch Gradient Descent

---

Cuando el algoritmo **Mini-Batch Gradient Descent** calcula las **derivadas acumuladas** se agrega al proceso el algoritmo **Exponentially Weighted Moving Average**, en conjunto logran minimizar el **Error** de la **Función de Error** mucho **mas rápido**

# RMSprop

Root Mean Square Propagation

---

## Algoritmo de Optimización para el algoritmo Mini-Batch Gradient Descent

---

En ves de **Acumular Derivadas**, el algoritmo **RMSprop** eleva al cuadrado el **Error** de la **Función** de **Error** y lo **normaliza** calculando raíces

# Adaptive Moment Optimization

ADAM

---



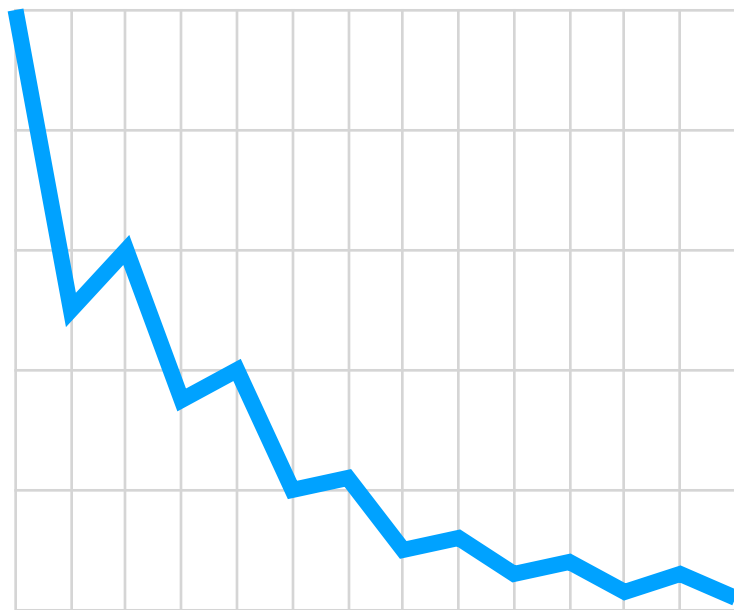
## Algoritmo de Optimización para el algoritmo Mini-Batch Gradient Descent

---

Combina ambos algoritmos para minimizar el **Error** de la **Función de Error**, utiliza **Gradient Descent with Momentum** para hacerlo rápido, y utiliza **RMSprop** para hacerlo mediante la normalización calculando raíces

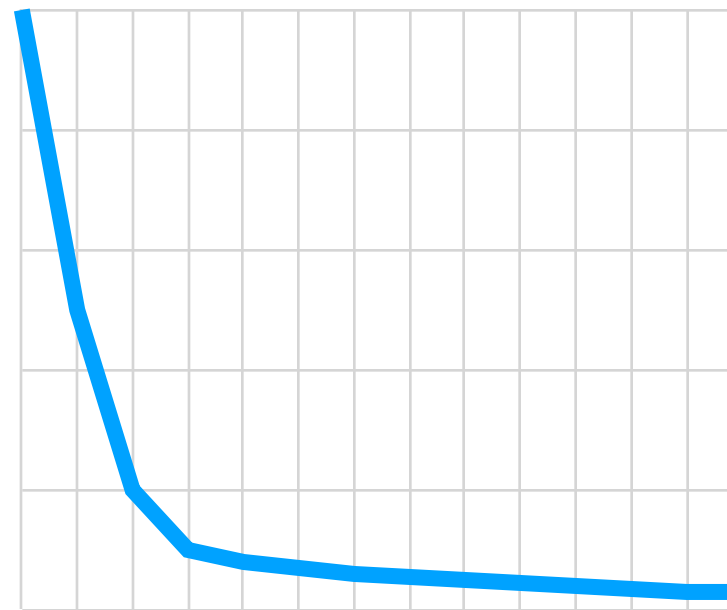
# Comparative

Mini-Batch  
Gradient Descent



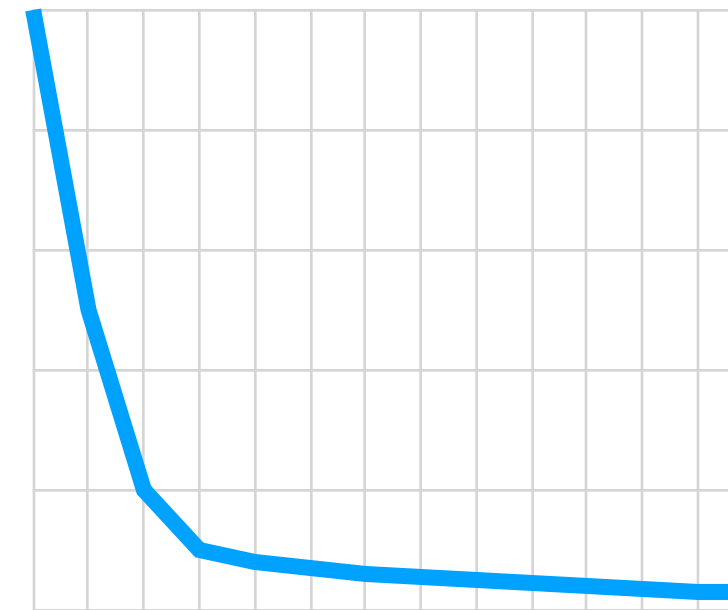
Minimización  
lenta del Error con  
Acumulación de  
Derivadas

Gradient Descent  
with Momentum



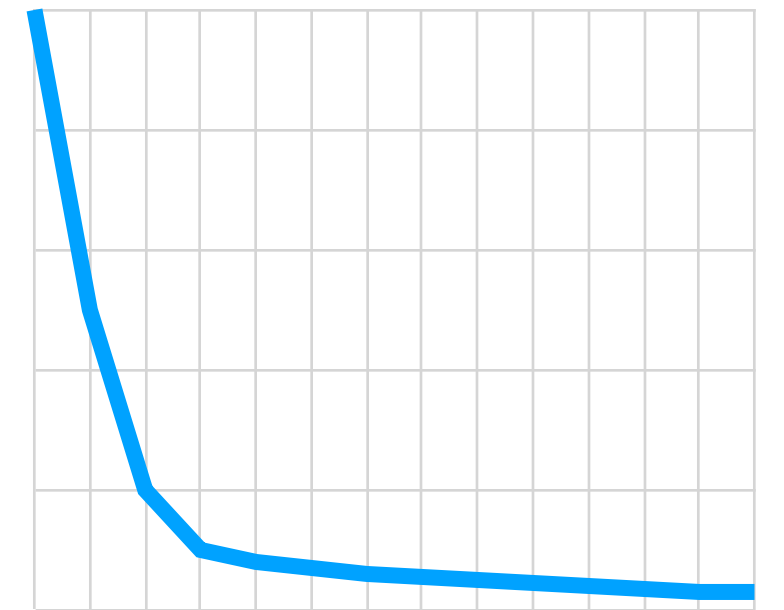
Minimización  
rápida del Error  
con Acumulación  
de Derivadas

RMSprop



Minimización  
rápida del Error  
con calculo de  
cuadrados y raíces

Adam



Minimización  
rápida del Error con:

Gradient Descent  
with Momentum

+

RMSprop

Todos se basan en Mini-Batch Gradient Descent