

# Métodos de optimización de la gradiente de descenso en una red neuronal convolucional

Víctor Jesús Sotelo Chico<sup>1</sup>

<sup>1</sup>Universidad Nacional de Ingeniería

Seminario de Tesis I



# Contenido

- 1 Introducción
- 2 Objetivos
- 3 Marco Teórico
  - Aprendizaje Automático
  - Redes Neuronales
- 4 Métodos de Optimización
  - Momentum
  - Nesterov
  - Adagrad
  - RMSprop
  - Adam
- 5 Resultados
- 6 Conclusiones y Trabajos Futuros



# Introducción

En la actualidad es indispensable emplear mucho tiempo en el entrenamiento de redes neuronales profundas, por lo que surge la necesidad de encontrar métodos que aceleren este proceso.



# Objetivos

- Entender las ventajas y desventajas de distintos métodos de optimización de la gradiente de descenso.
- Obtener la capacidad de discriminar entre distintos métodos de optimización.
- Lograr un mejor entendimiento de las redes neuronales profundas.



# Aprendizaje Automático

Se encarga consiste aprenden a identificar patrones en un conjunto de datos. A medida que se realiza este aprendizaje, la máquina podrá ser capaz de realizar una predicción o tomar decisiones sin haber estado programada explícitamente para realizar esta tarea.

El Aprendizaje Automático puede ser dividido de la siguiente forma :

- Aprendizaje Supervisado
- Aprendizaje No Supervisado
- Aprendizaje por Refuerzo



# Aprendizaje Supervisado

- Regresión Lineal
- Regresión Logística
- Clasificación



# Aprendizaje No Supervisado





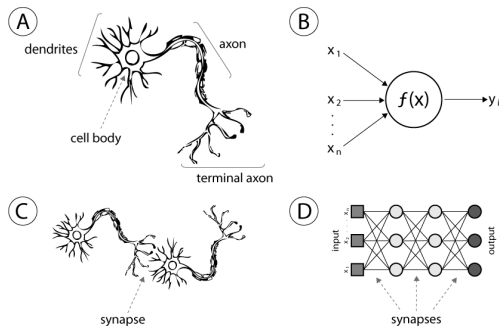


# Redes Neuronales Artificiales

Estas redes toman como inspiración la arquitectura del cerebro para la construcción de sistemas inteligente. Actualmente son la base para el desarrollo de la inteligencia artificial.



# Comparación neuronas biológicas y artificiales



**FIGURE –** Redes neuronales biológicas y artificiales

# Redes neuronales Prealimentadas

Es un tipo de red neuronal más simple que existe. Esta red puede clasificarse en :

- Perceptron simple
- Perceptron Multicapas
- Redes neuronales convolucionales



# Esquema Redes neuronales Prealimentadas

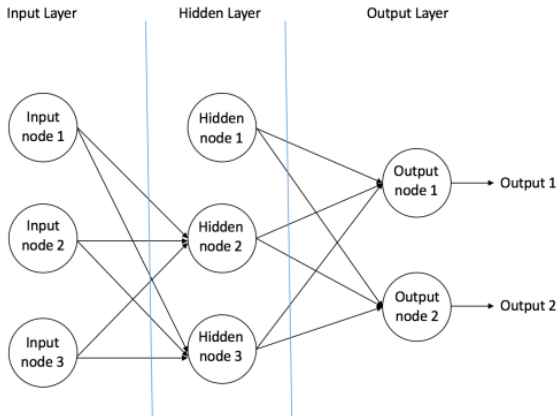


FIGURE – Esquema de Redes Neuronales Prealimentadas

# Back Propagation



Introducción

Objetivos

**Marco Teórico**

Métodos de Optimización

Resultados

Conclusiones y Trabajos Futuros

Aprendizaje Automático

Redes Neuronales

# Redes Neuronales Convolucionales



# Capas de una red neuronal convolucional

- Input Layer
- Convolutional Layer
- Pooling Layer
- Fully Connected Layer
- Output Layer



# Gradiente de Descenso





# Variantes de la Gradiente de Descenso

Existen 3 variantes de la gradiente de descenso :

- Batch gradient descent
- Stochastic gradient descent
- Mini-batch gradient descent



# Métodos para optimizar la gradiente de descenso

- Momentum
- Nesterov Momentum
- Adagrad
- RMSprop
- Adam



# Momentum

$$\begin{aligned}\nu_t &= \gamma \nu_{t-1} + \eta \nabla_{\theta} J(\theta) \\ \theta &= \theta - \nu_t\end{aligned}\tag{1}$$



# Nesterov

$$\begin{aligned}\nu_t &= \gamma \nu_{t-1} + \eta \nabla_{\theta} J(\theta - \gamma \nu_{t-1}) \\ \theta &= \theta - \nu_t\end{aligned}\tag{2}$$



# Adagrad

$$\begin{aligned} g_{t,i} &= \nabla_{\theta} J(\theta_{t,i}) \\ \theta_{t+1,i} &= \theta_{t,i} - \eta \cdot g_{t,i} \end{aligned} \tag{3}$$

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} \cdot g_{t,i} \tag{4}$$



# RMSprop

$$\begin{aligned} E[g^2]_t &= \gamma E[g^2]_{t-1} + (1 - \gamma) g_t^2 \\ \theta_{t+1} &= \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t \end{aligned} \tag{5}$$



# Adam

$$\begin{aligned}m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2\end{aligned}\tag{6}$$



# Resultados

Para obtener nuestros resultados utilizamos 2 datasets :

- CIFAR 10
- CIFAR 100





# Resultados CIFAR-10



# Resultados CIFAR-100



# Conclusiones



# Trabajos Futuro

