



Universidad Michoacana de San Nicolás de Hidalgo
Instituto de Física y Matemáticas



Aplicación de redes neuronales convolucionales a la discriminación gamma-hadrón en chubascos atmosféricos

Presenta: Armando Madrigal Lucatero

Director de tesis:

Dr. José Antonio González Cervera

Codirector de tesis:

Dr. Cederik León De León Acuña

30 de Agosto de 2021

Motivacion:

Motivacion:

- Los observatorios de rayos gamma como HAWC son capaces de detectar chubascos atmosfericos inducidos tanto por rayos gamma como otros tipos de rayos cosmicos.
- Para separar las señales de rayos gamma es necesario la implemetación de metodos que permitan filtrar estas señales.

Objetivos:

- Estudiar la técnica de redes neuronales artificiales (RNA) para aplicarlas a la solución del problema.
- Implementar un método de separación gamma-hadrón capaz de resolver el problema de forma eficiente.

Chubascos Atmosféricos

- Los rayos cósmicos (RC) están compuestos por diferentes tipos de partículas procedentes del espacio exterior cuya energía es muy elevada (superior a los 0.3 GeV).
- Los RC están compuestos principalmente por protones (86%), partículas alfa (11%), electrones (2%) y núcleos pesados como el uranio (1%).
- Cuando los RC interactúan con la atmósfera terrestre se producen chubascos atmosféricos
- Los chubascos atmosféricos están constituidos por diferentes tipos de procesos muy complejos, los cuales involucran interacciones fundamentales de la física de partículas, la física de RC y procesos astrofísicos.



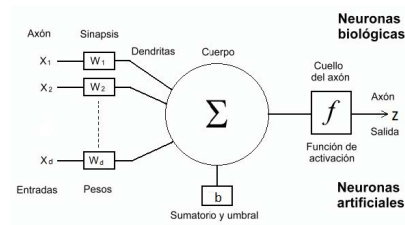
Rayos gamma

- En la astronomía y la astrofísica se estudian los rayos gamma como herramienta para diversos fines, por ejemplo, estudiar explosiones de supernova, descubrir nuevos objetos astrofísicos, estudiar agujeros negros, estudiar la materia en condiciones extremas para la búsqueda de nueva física.
- El número de eventos inducidos por hadrones que se registran, es varios ordenes de magnitud más grande que los eventos inducidos por rayos gamma.



Redes neuronales artificiales

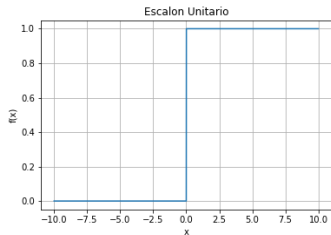
- Las redes neuronales artificiales (RNA) son un modelo computacional de aprendizaje automático.
- Inspiradas en el funcionamiento del cerebro humano
- La unidad básica de una RNA es la neurona artificial



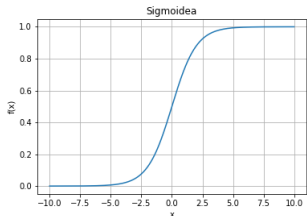
$$z = f \left(\sum_{i=1}^d w_i x_i + b \right), \quad (1)$$

Redes neuronales artificiales

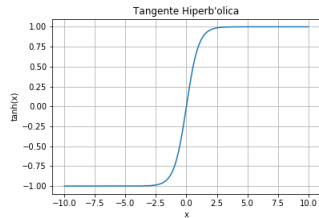
- Función escalón unitario



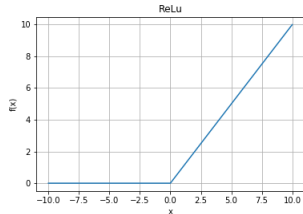
- Función sigmoide



- Función tangente hiperbólica



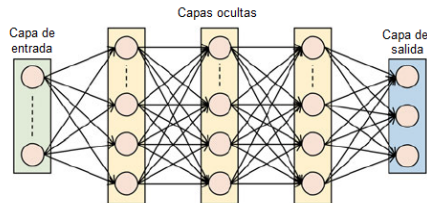
- Función Relu



Redes neuronales artificiales

$$z_i^{(c)} = f \left(\sum_{j=1}^{N_c} w_{ij}^{(c)} x_j^{(c)} + b_i^{(c)} \right), \quad (2)$$

Perceptron multicapa

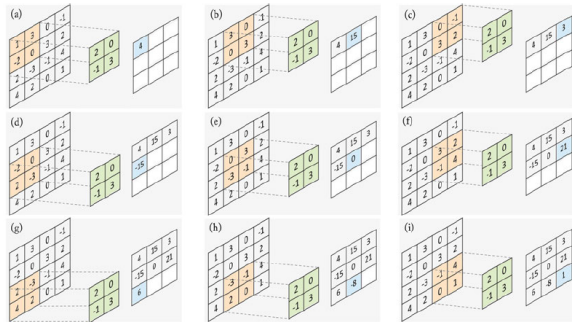


Redes neuronales convolucionales

Se define la operación de convolución como sigue:

$$(X * W)_{r,s} = M_{r,s} = \sum_{i=1}^k \sum_{j=1}^l X_{r+i-1,s+j-1} W_{i,j}, \quad (3)$$

donde X corresponde a una matriz de características de tamaño m por n y W es un filtro de tamaño k por l , con $k < m$ y $l < n$, además $r = 1, 2, \dots, m - k + 1$, $s = 1, 2, \dots, n - l + 1$.



Redes neuronales convolucionales

Capa convolucional

Supongamos que tenemos un objeto \mathbf{X} que representa un mapa bidimensional compuesto por I canales, $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(I)}$, este mapa es la entrada de una capa de convolución, la cual nos da como salida otro mapa bidimensional \mathbf{Z} de J canales, $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \dots, \mathbf{Z}^{(J)}$. La operación que realiza una capa de convolución es la siguiente:

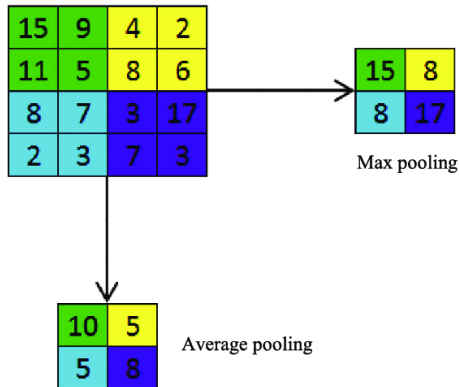
$$\mathbf{Z}^{(j)} = f \left(\sum_{i=1}^I \mathbf{X}^{(i)} * \mathbf{W}^{(j)} + B_j \right), \quad (4)$$

donde $\mathbf{W}^{(j)}$ es el j -ésimo filtro de la convolución y B_j es un parámetro de sesgo, con $j = 1, 2, \dots, J$, donde J es el número de filtros de la convolución, además, $f(t)$ es una función de activación.

Redes neuronales convolucionales

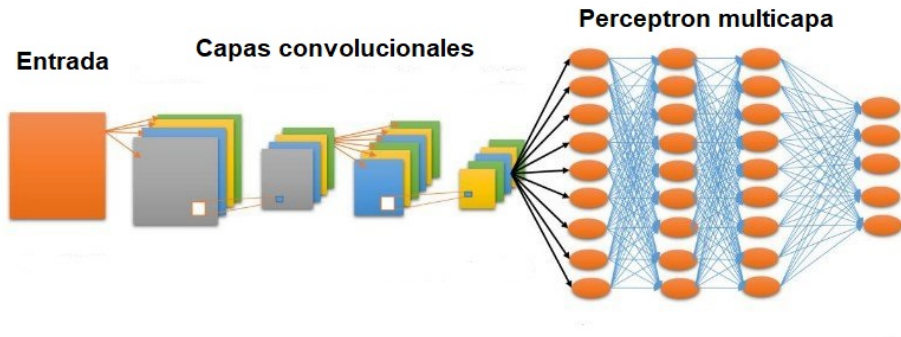
Reducción de muestreo

- *Max pooling*: Extrae el elemento máximo.
- *Average pooling*: Extrae el promedio.



Redes neuronales convolucionales

Esquema gráfico de una red neuronal convolucional



Aprendizaje supervisado

- Consiste en entrenar un modelo de aprendizaje automático con datos previamente etiquetados.
- El modelo aprende a generalizar el problema y es capaz de predecir nuevos resultados.

Se utilizan etiquetas vectoriales para etiquetar los datos:

$$\text{clase } 1 : (1, 0, \dots, 0), \quad (5)$$

$$\text{clase } 2 : (0, 1, \dots, 0), \quad (6)$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad (7)$$

$$\text{clase } N_y : (0, 0, \dots, 1). \quad (8)$$

Entrenamiento de redes neuronales

- Consiste en buscar los parámetros de la red neuronal, que mejor se ajusten a los datos de entrenamiento.
- Capa de salida *SoftMax*

$$\hat{y}_j = P_j(\mathbf{z}) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (9)$$

donde \hat{y}_j es la salida de la j-ésima neurona de la capa de salida, además

$$z_j = z_j^{(C)} = \sum_{k=1}^{N_C} w_{jk}^{(C)} x_k^{(C)} + b_j^{(C)}, \quad (10)$$

- Se define la función de error entropía cruzada categorica:

$$e(\mathbf{w}) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{N_C} y_{ij} \log(P_{ij}) \quad (11)$$

donde n es el número de ejemplos de entrenamiento en cada lote, y_{ij} es la salida deseada para el i-ésimo ejemplo en la j-ésima coordenada del vector de salida

Entrenamiento de redes neuronales

Descenso de gradiente

$$w_i^{(t+1)} = w_i^{(t)} - \eta \frac{\partial e(\mathbf{w})}{\partial w_i}, \quad (12)$$

- Descenso de gradiente estocástico: consiste en ajustar el error para un subconjunto del total de datos de entrenamiento elegidos de forma aleatoria.

Se dividen los datos en tres conjuntos:

- Datos de entrenamiento.
- Datos de validación.
- Datos de prueba.

Datos de entrenamiento	Datos de validación	Datos de prueba
60%	20%	20%

Simulación de chubascos atmosféricos

- Simulador de rayos cósmicos CORSIKA (COsmic Ray Simulations for KAscade), basado en el método de Monte Carlo.
- Consideramos solo chubascos verticales

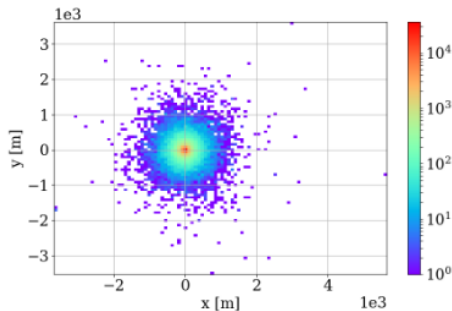
Partícula primaria: partícula (gamma o hadrón) que induce el chubasco atmosférico.

Partículas secundarias: partículas medidas al nivel del piso producidas por el chubasco atmosférico simulado.

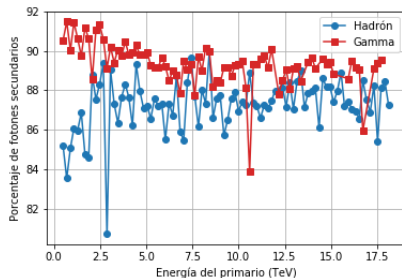
ID Partícula	$P_x[\text{eV}]$	$P_y[\text{eV}]$	$P_z[\text{eV}]$	$x[\text{m}]$	$y[\text{m}]$
1	0.0013842	0.00619146	0.0181968	36540.9	202339
3	-0.000463009	0.00170897	0.0013097	21935.1	-6366.1
13	-0.201099	-0.108235	1.21352	-103682	-56964.4
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
2	0.00212595	-0.00754431	0.0281506	-2382.11	-660.024

Simulación de chubascos atmosféricos

Densidad de partículas secundarias medidas en el piso para un chubasco inducido por una partícula gamma con energía de 19.51 TeV.



Porcentaje de fotones en las partículas secundarias para chubascos inducidos por partículas gamma y hadrones con diferente energía.



Simulación de chubascos atmosféricos

Mapas bidimensionales de tamaño 100 por 100 con 4 canales:

- **Canal 1:** contiene información de la distribución de fotones en la huella al piso.
- **Canal 2:** contiene información del promedio del momento lineal de los fotones en cada región de la huella al piso.
- **Canal 3:** contiene información de la distribución de no-fotones en la huella al piso.
- **Canal 4:** contiene información del promedio del momento lineal de los no-fotones en cada región de la huella al piso.

Simulación de chubascos atmosféricos

Decimos que una partícula cayó en la celda (i, j) , si para su posición (x, y) se cumple que,

$$x_{min} + (j - 1)\Delta x \leq x \leq x_{min} + j\Delta x \quad (13)$$

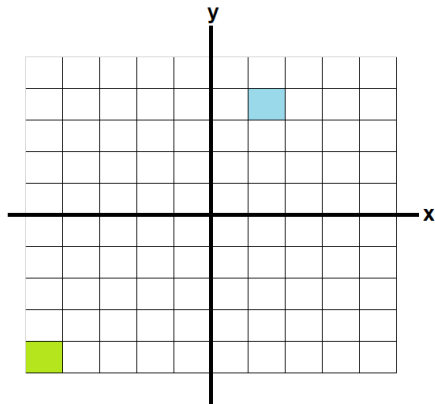
$$y_{max} - i\Delta y \leq y \leq y_{max} - (i - 1)\Delta y, \quad (14)$$

con

$$\Delta x = \frac{x_{max} - x_{min}}{100} \quad (15)$$

y

$$\Delta y = \frac{y_{max} - y_{min}}{100}, \quad (16)$$



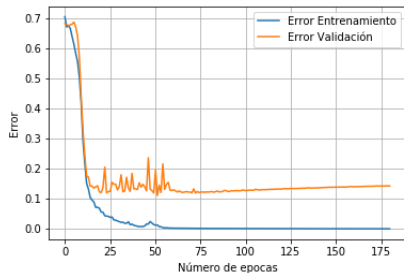
Clasificación gamma-hadrón

- 527 simulaciones de partículas gamma y hadrones con energía entre 0.5 y 30 TeV.
- Se entrenó una red neuronal convolucional durante 180 épocas con 11 pasos por época, con lotes de 30 datos en cada paso.
- La estructura de la red se muestra en la siguiente tabla:

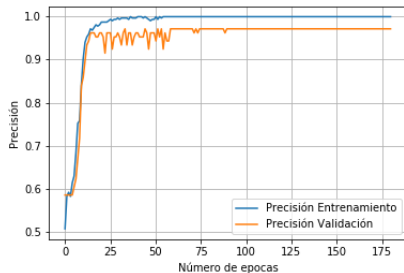
Número de capa:	Tipo de capa:	Información de los hiperparámetros:
1	Capa Convolucional	20 filtros de tamaño 5x5, función de activación <i>ReLU</i>
2	Capa MaxPooling	ventana <i>MaxPooling</i> de tamaño 2x2
3	Capa Convolucional	10 filtros de tamaño 4x4, función de activación <i>ReLU</i>
4	Capa MaxPooling	ventana de <i>MaxPooling</i> de tamaño 2x2
5	Capa Densa	80 neuronas, función de activación <i>ReLU</i>
6	Capa Densa	40 neuronas, función de activación <i>sigmoide</i>
7	Capa Densa	2 neuronas, función de activación <i>softmax</i>

Clasificación gamma-hadrón

Resultados:



- Error de validación mínimo en época 52

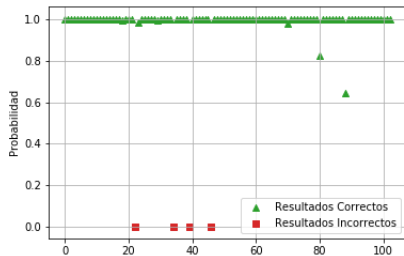


- Máxima precisión de validación: 97%

Clasificación gamma-hadrón

Resultados:

- Datos de prueba



- Eficiencia del modelo: 99/104 (96%)

Clasificación en intervalos de energía

- 2680 simulaciones de chubascos atmosféricos inducidos por partículas gamma y hadrones con energías entre 0.5 y 15 TeV.
- Se realizaron 3 pruebas con diferentes estructuras para la red neuronal.
- Se clasifican las simulaciones en 6 diferentes categorías.

Número de clase:	Información:	Etiqueta vectorial:
1	Gamma 0.5 - 5.0 TeV	(1,0,0,0,0,0)
2	Gamma 5.0 - 10.0 TeV	(0,1,0,0,0,0)
3	Gamma 10.0 - 15.0 TeV	(0,0,1,0,0,0)
4	Hadrón 0.5 - 5.0 TeV	(0,0,0,1,0,0)
5	Hadrón 5.0 - 10.0 TeV	(0,0,0,0,1,0)
6	Hadrón 10.0 - 15.0 TeV	(0,0,0,0,0,1)

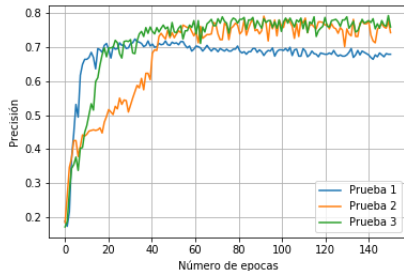
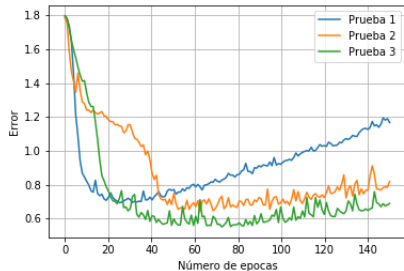
Clasificación en intervalos de energía

Prueba:	Profundidad de la RNC:	Número de parámetros de la RNC:
1	7 Capas	395,996
2	10 Capas	95,846
3	13 Capas	28,006

Table: Se compara el numero de parámetros de cada prueba realizada. Se muestra como este disminuye al aumentar la profundidad de la RNC.

Clasificación en intervalos de energía

Resultados:

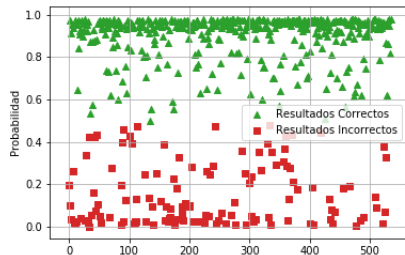


Prueba:	Época óptima:	Precisión de entrenamiento:	Precisión de validación:
1	32	82%	71%
2	63	84%	76%
3	73	82%	78%

Clasificación en intervalos de energía

Resultados:

- Datos de prueba:



- Eficiencia del modelo: 411/536 (76%)

Conclusiones

- Se mostró que el modelo propuesto tiene una alta eficiencia (96%) en la separación gamma-hadrón.
- El modelo de redes neuronales mejora su eficiencia al aumentar la profundidad en el número de capas
- Aunque el modelo propuesto tiene una alta eficiencia en la separación gamma-hadrón, esta eficiencia baja del 96% al 76% cuando se considera la clasificación de la energía.

Trabajo futuro

- Mejorar las técnicas de separación gamma-hadrón implementando otros modelos de inteligencia artificial distintos a las redes neuronales
- Considerar más variables en las simulaciones de chubascos atmosféricos como, diferentes inclinaciones y diferentes tipos de atmósfera.