



Aprendizaje automático

Ricardo Balderas, Oscar Chaparro,
Alberto Maldonado, Miguel Martínez.
Jesús Alberto Martínez Castro.

22 de noviembre de 2020

Instituto Politécnico Nacional
Centro de Investigación en Computación

Agenda

Introducción

Conceptos

La representación del conocimiento

Aplicaciones

Referencias

Introducción



Conceptos



La representación del conocimiento



Aplicaciones



Referencias



En Dios confiamos,
todos los demás tenemos que
aportar **datos**.

William Deming.

En Dios confiamos,
todos los demás tenemos que
aportar **datos**.

William Deming.

Los **datos** son el nuevo crudo.

Clive Humby.

Dato proviene del latin “datum” cualquier cosa o hecho dado por cierto.

Dato proviene del latin “**datum**” cualquier cosa o hecho dado por cierto.

En la actualidad “**Dato**” es cualquier representación simbólica que se pueda almacenar, analizar y reorganizar.

Primer experimento de recolección de **datos** para predecir sucesos posteriores fue en **1663**. El estadístico inglés John Graunt recopiló información sobre la **mortalidad en Londres** y gracias a ello pudo **alertar** con antelación del **resurgimiento** de la **peste bubónica** en Europa.

En **1939** empleados de una empresa de transportes de Londres examinaron más de **cuatro millones de billetes** usados de metro.



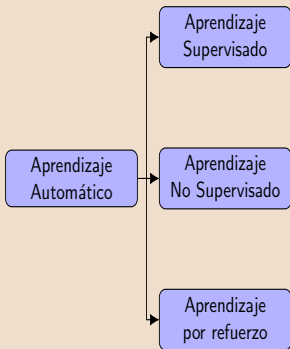
Primer experimento de recolección de **datos** para predecir sucesos posteriores fue en **1663**. El estadístico inglés John Graunt recopiló información sobre la **mortalidad en Londres** y gracias a ello pudo **alertar** con antelación del **resurgimiento** de la **peste bubónica** en Europa.

En **1939** empleados de una empresa de transportes de Londres examinaron más de **cuatro millones de billetes** usados de metro.

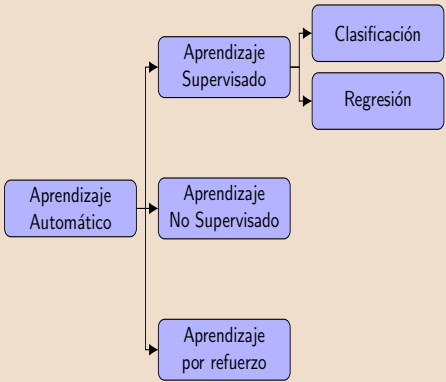


Se buscaba **encontrar** cuáles eran las rutas más o menos **utilizadas** y así poder ayudar a desarrollar **nuevas infraestructuras**.

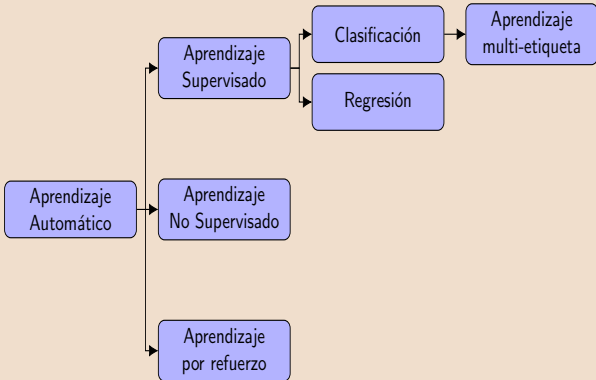
Ramas principales del aprendizaje automático



Ramas principales del aprendizaje automático

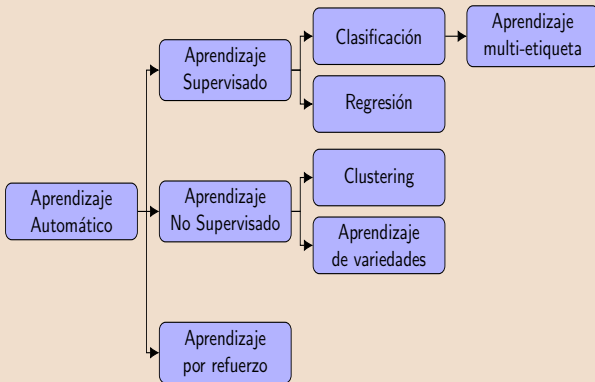


Ramas principales del aprendizaje automático





Ramas principales del aprendizaje automático





Aprendizaje supervisado

Aprendizaje supervisado

El nombre de aprendizaje supervisado proviene del hecho que es **necesaria** la **intervención** de un maestro o **supervisor** que **asigne** correctamente el valor de la **etiqueta** en casos previamente adquiridos.

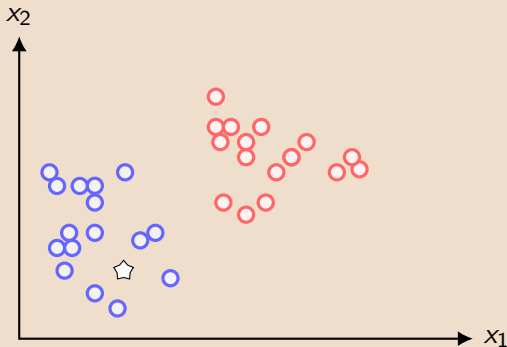
Aprendizaje supervisado

El nombre de aprendizaje supervisado proviene del hecho que es **necesaria** la **intervención** de un maestro o **supervisor** que **asigne** correctamente el valor de la **etiqueta** en casos previamente adquiridos.

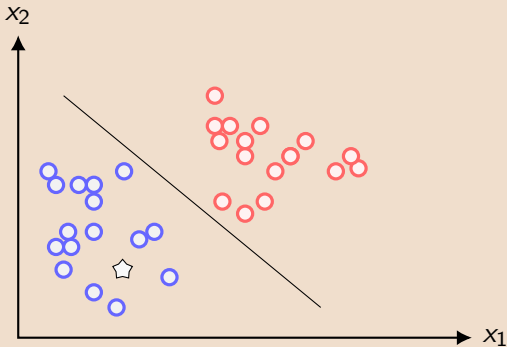
Desde un punto de vista formal, **dado** un conjunto de parejas **dato-etiqueta**, el aprendizaje supervisado **busca** un modelo que permita realizar una **correspondencia** de los datos a las etiquetas **de tal forma que** dado un **nuevo dato** que no se haya visto anteriormente **el algoritmo** de aprendizaje supervisado **permita predecir la etiqueta** que le correspondería.

Cuando la etiqueta puede tomar un **número finito** de valores estamos frente a un problema de **clasificación**.

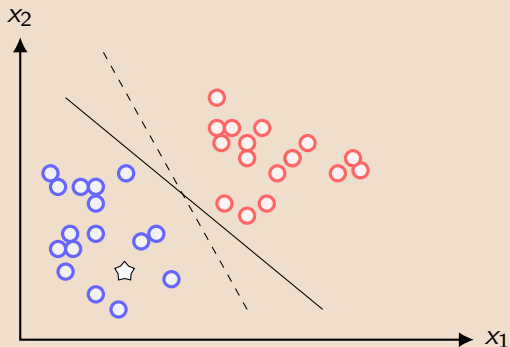
Cuando la etiqueta puede tomar un **número finito** de valores estamos frente a un problema de **clasificación**.



Cuando la etiqueta puede tomar un **número finito** de valores estamos frente a un problema de **clasificación**.



Cuando la etiqueta puede tomar un **número finito** de valores estamos frente a un problema de **clasificación**.



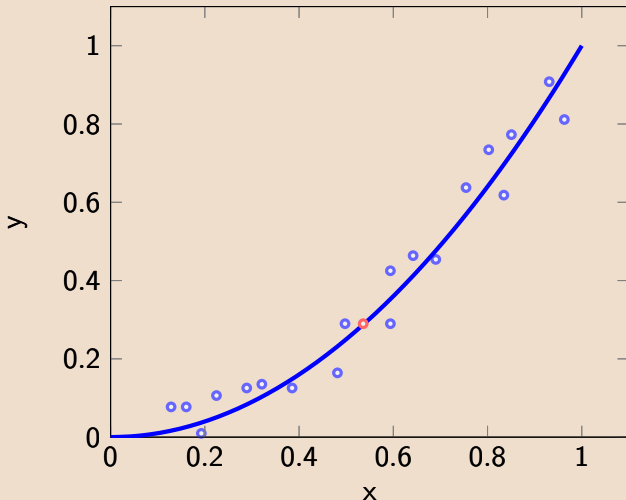
En ocasiones los datos pueden tener **más de una etiqueta**. Por ejemplo, en una misma película se pueden considerar varios géneros.



Figura: Aprendizaje Multi-etiqueta.

Si la etiqueta puede tomar un **número infinito** de valores, por ejemplo la predicción de un número real, llamamos a ese problema, **regresión**.

Si la etiqueta puede tomar un **número infinito** de valores, por ejemplo la predicción de un número real, llamamos a ese problema, **regresión**.



Aprendizaje no supervisado

Aprendizaje no supervisado

Este tipo de aprendizaje se distingue del supervisado por que **no existe etiqueta** que se quiera predecir, ni un tutor que anote previamente los datos e identifique cual es el resultado deseado.

El aprendizaje no supervisado es importante cuando **objetivo** del aprendizaje es **explorar** y ganar más información sobre **los datos**.

Aprendizaje no supervisado

Este tipo de aprendizaje se distingue del supervisado por que **no existe etiqueta** que se quiera predecir, ni un tutor que anote previamente los datos e identifique cual es el resultado deseado.

El aprendizaje no supervisado es importante cuando **objetivo** del aprendizaje es **explorar** y ganar más información sobre **los datos**.

Su componente clave es el concepto de **similitud entre los datos**.

Aprendizaje no supervisado

Este tipo de aprendizaje se distingue del supervisado por que **no existe etiqueta** que se quiera predecir, ni un tutor que anote previamente los datos e identifique cual es el resultado deseado.

El aprendizaje no supervisado es importante cuando **objetivo** del aprendizaje es **explorar** y ganar más información sobre **los datos**.

Su componente clave es el concepto de **similitud entre los datos**.

Se pueden distinguir **dos** grandes **sub-familias** dentro de este grupo:

Métodos de agrupación o mixturas o clustering

Pretende dividir los datos en **grupos excluyentes** (un dato en un solo grupo) o **no excluyentes** (un dato puede pertenecer a diversos grupos con distinto índice de pertenencia).

Métodos de agrupación o mixturas o clustering

Pretende dividir los datos en grupos excluyentes (un dato en un solo grupo) o no excluyentes (un dato puede pertenecer a diversos grupos con distinto índice de pertenencia).

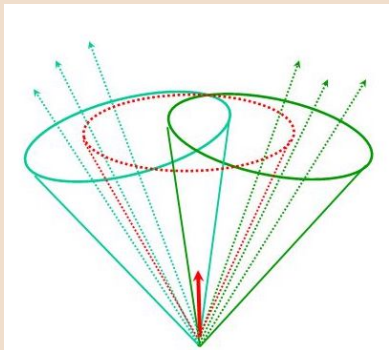


Figura: Una partícula puede o no pertenecer a jet dependiendo de parámetros geométricos.

Introducción
○○○

Conceptos
○○○○○○○○●○

La representación del conocimiento
○○○○○○○○○○○○○○

Aplicaciones
○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○

Referencias
○

Aprendizaje de variedades

Desde un punto de vista informal e intuitivo, una variedad es un **espacio** tal que nuestra **intuición** sobre lo que está cerca y lejos **se mantiene**.

Datos sin ordenar (ninguna caricia previa)

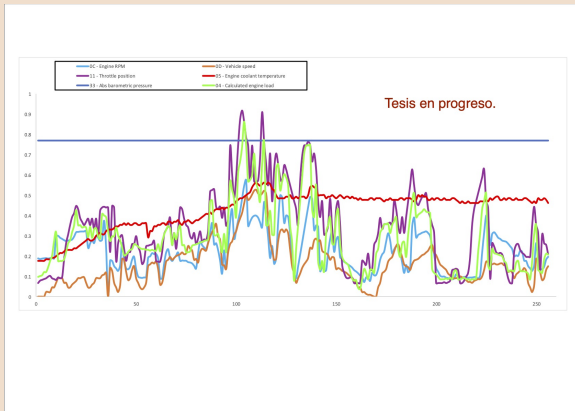


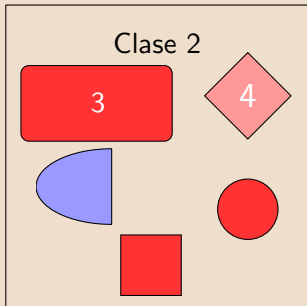
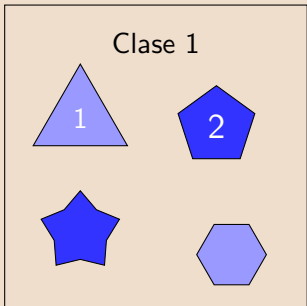
Figura: Diversos parámetros de conducción ¹.

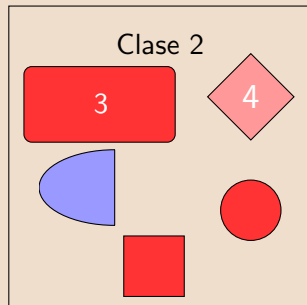
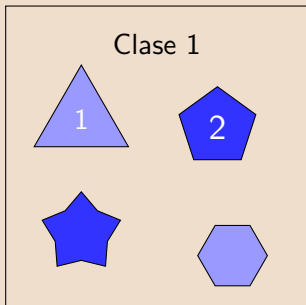
¹Gracias a Raúl Castillo Luna por permitirnos mostrar sus gráficas

El **primer paso** necesario para el diseño y uso de los algoritmos de aprendizaje automático es la **representación del conocimiento** o de la experiencia previa.

Para ello se necesita, en general, una **representación numérica** de los **conceptos** y de las **abstracciones** de la vida real.

A este proceso se le conoce como **extracción de características** y consiste en la **descripción de rasgos** relevantes de los objetos de interés **en relación a la tarea** que debe llevarse a cabo y su posterior representación numérica.





Elemento	Color	Número de vértices	Vértices redondos	Clase
1	Azul claro	3	no	1
2	Azul oscuro	5	no	1
3	Rojo oscuro	4	sí	2
4	Rojo claro	4	no	2

Introducción
○○○

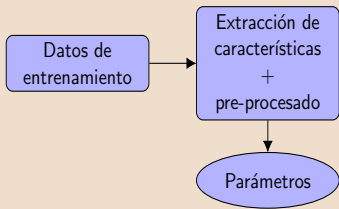
Conceptos
○○○○○○○○○○○

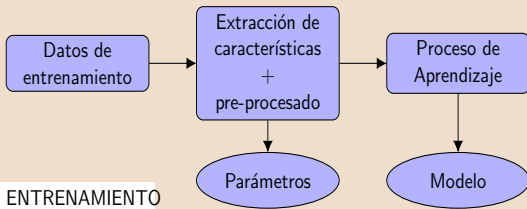
La representación del conocimiento
○○○○●○○○○○○○○○

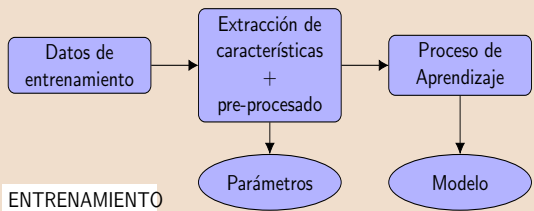
Aplicaciones
○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○

Referencias
○

Datos de
entrenamiento

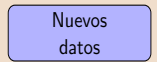


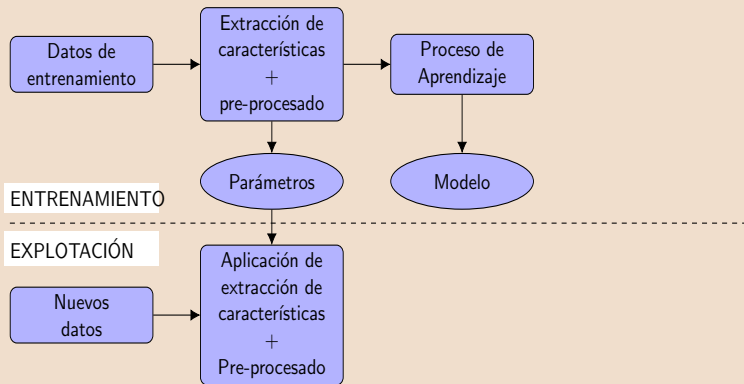


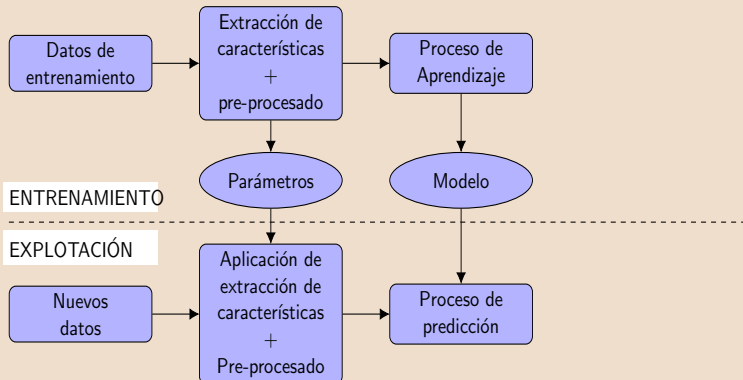


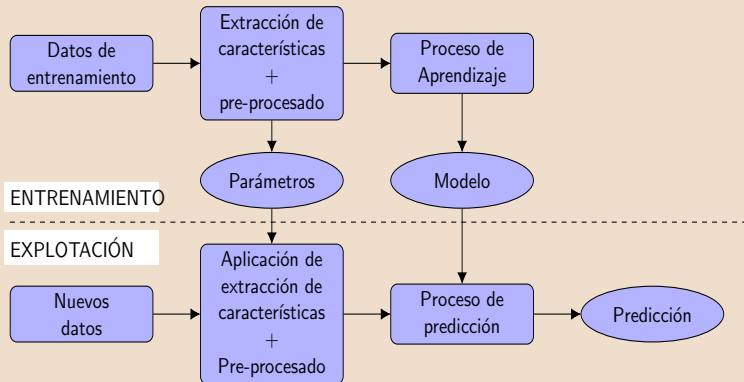
ENTRENAMIENTO

EXPLOTACIÓN









Métodos usados en Aprendizaje Automático

Regresión lineal[3]

Pros: Resultados fáciles de interpretar, computacionalmente barato.

Cons: No es muy bueno modelando información no lineal.

Usar con: Valores numéricos y nominales.

Agrupamiento por k-medias (kNN) [3]

Pros: Fácil de implementar.

Cons: Puede converger a un mínimo local, lento en conjuntos grandes.

Usar con: Valores numéricos.

Máquinas de vectores de soporte

Pros: Error de generalización bajo, computacionalmente barato y fácil de interpretar.

Cons: Sensible a ajustes en los parámetros y al kernel elegido. De forma nativa, solo trabaja con clasificación binaria.

Usar con: Valores numéricos y nominales.

Regresión basada en árboles[3]

Pros: Se ajusta a información compleja y no lineal.

Cons: Los resultados son difíciles de interpretar.

Usar con: Valores numéricos y nominales.

Redes neuronales

Es un sistema capaz de **aprender** las **mejores características** para un determinado problema, y es la **base de** la rama de aprendizaje automático conocida como **aprendizaje profundo**.

Perceptrón

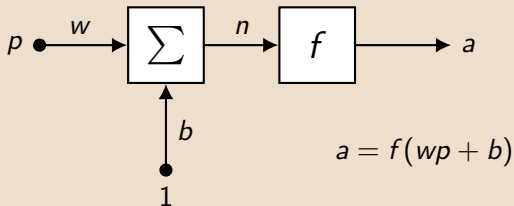


Figura: Una neurona.

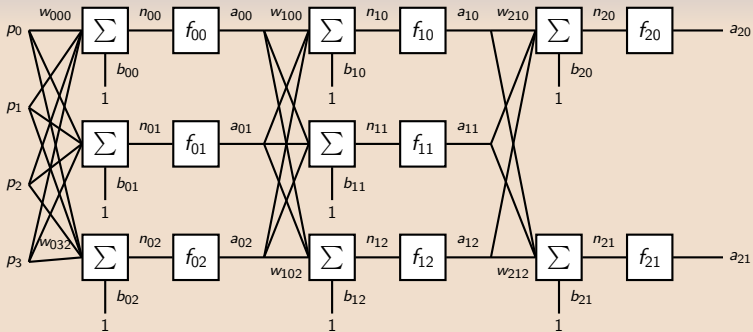


Figura: Perceptrón multicapa.

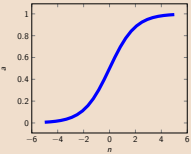
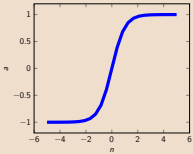
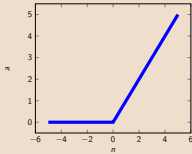
$$a_0 = f_0(W_0 p + b_0), \quad a_1 = f_1(W_1 a_0 + b_1), \quad a_2 = f_2(W_2 a_1 + b_2)$$

$$p = \begin{pmatrix} p_0 \\ \vdots \\ p_r \end{pmatrix}, \quad a_i = \begin{pmatrix} a_{i0} \\ \vdots \\ a_{i_r} \end{pmatrix}, \quad W_i = \begin{pmatrix} w_{i00} & \cdots & w_{i0c} \\ \vdots & \ddots & \vdots \\ w_{ir0} & \cdots & w_{irc} \end{pmatrix}, \quad b_i = \begin{pmatrix} b_{i0} \\ \vdots \\ b_{i_r} \end{pmatrix}$$

Funciones de activación

¿Por qué? Incorporan la no linealidad que le da poder a una red neuronal.

¿Cuáles? Existen muchas, pero las que usaremos son:

Sigmoide	Tangente hiperbólica	ReLU
$1/(1 + e^{-n})$	$\tanh(n)$	$\max(0, n)$
		

Más información en...

- Peter Harrington. *Machine Learning in Action*.
Manning Publications Co., USA, 2012
- Howard B. Demuth, Mark H. Beale, Orlando de Jesús, and
Martin T. Hagan. *Neural Network Design*.
Martin Hagan, Stillwater, OK, USA, 2nd Edition, 2014

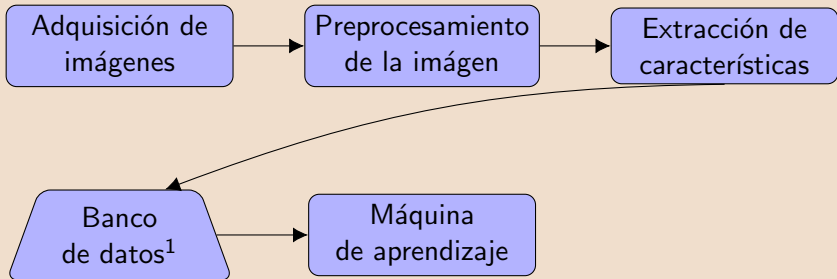
Los ejercicios contemplados para el día de mañana son:

- Clasificación de imágenes usando Keras (redes neuronales).
- Entrenamiento supervisado usando Tensorflow.
Requerimientos.

Ejemplo 1

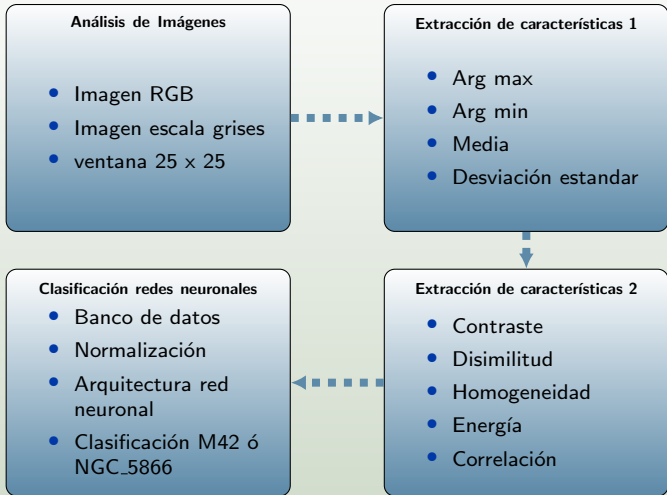
Análisis de imágenes y Aprendizaje automático

Diagrama para clasificar fragmentos de imágenes que pertenecen a una nebulosa o a una galaxia

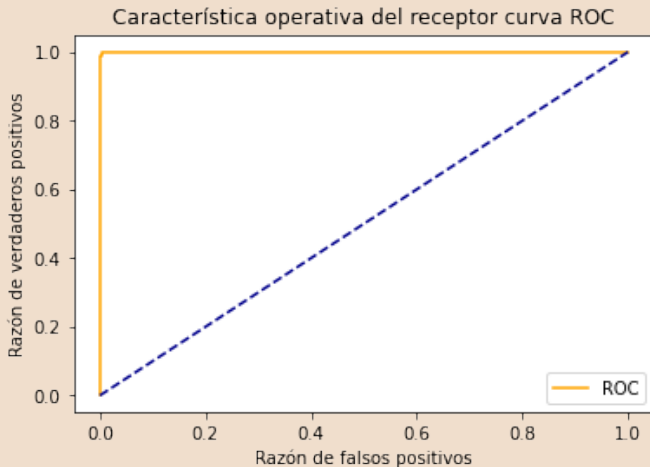


Análisis de imágenes y Aprendizaje automático

Diagrama técnico para clasificación de imágenes usando aprendizaje supervisado



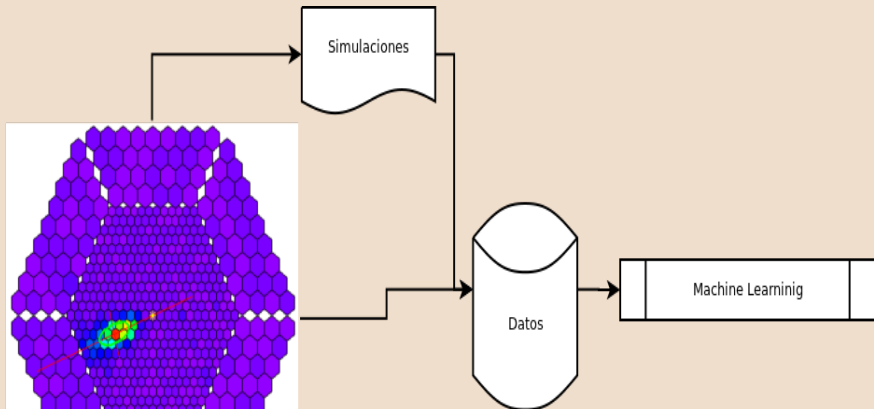
Curva ROC¹



¹ROC: Receiver Operating Characteristic

Ejemplo 2

Predicción de energía

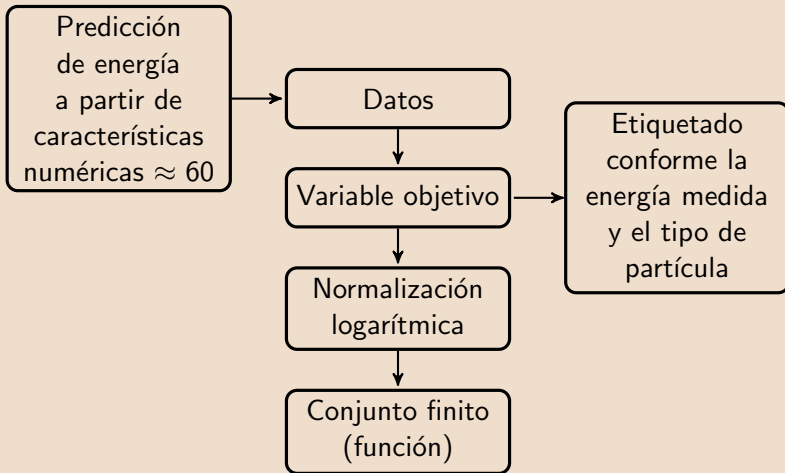


0.60°
189mm

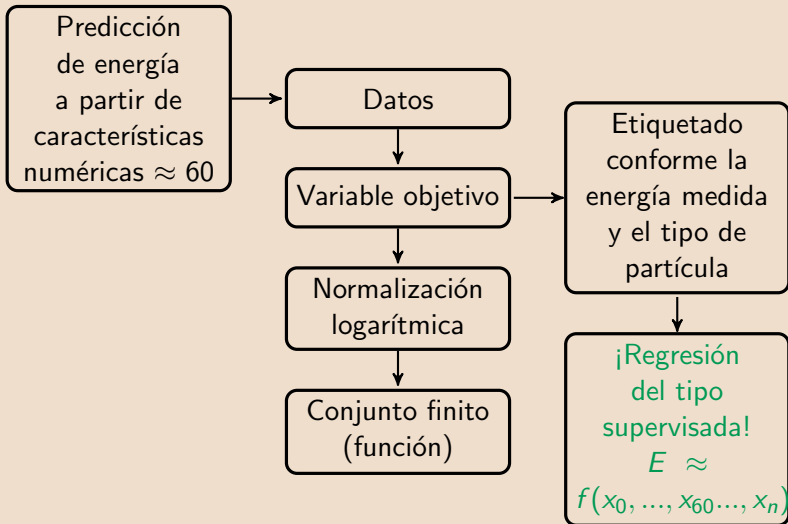
Formas del impacto de energía tanto de rayos gammas como de hadrones.

- * Valores de energía finito.
- * Set de características no nulas.
- * No se cuenta con formas explícitas con suficiente exactitud de predicción de la energía detectada en función de sus características.

Problema particular



Problema particular



Modelos de predicción y sus métricas

- Ciertas características son más importantes que otras. ¿Alguna característica se puede descartar?,
¿ Valdría la pena?

Modelos de predicción y sus métricas

- Ciertas características son más importantes que otras. ¿Alguna característica se puede descartar?,
¿ Valdría la pena? **Sí (proceso de descubrimiento).**

Modelos de predicción y sus métricas

- Ciertas características son más importantes que otras. ¿Alguna característica se puede descartar?,
¿ Valdría la pena? **Sí (proceso de descubrimiento).**
- ¿La importancia de las características depende del algoritmo empleado?

Modelos de predicción y sus métricas

- Ciertas características son más importantes que otras. ¿Alguna característica se puede descartar?,
¿ Valdría la pena? **Sí (proceso de descubrimiento).**
- ¿La importancia de las características depende del algoritmo empleado?
Sí (depende de los patrones encontrados).

Modelos de predicción y sus métricas

- Ciertas características son más importantes que otras. ¿Alguna característica se puede descartar?,
¿ Valdría la pena? **Sí (proceso de descubrimiento)**.
- ¿La importancia de las características depende del algoritmo empleado?
Sí (depende de los patrones encontrados).

- Factor de correlación  ≈ 1  ≈ 0 :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

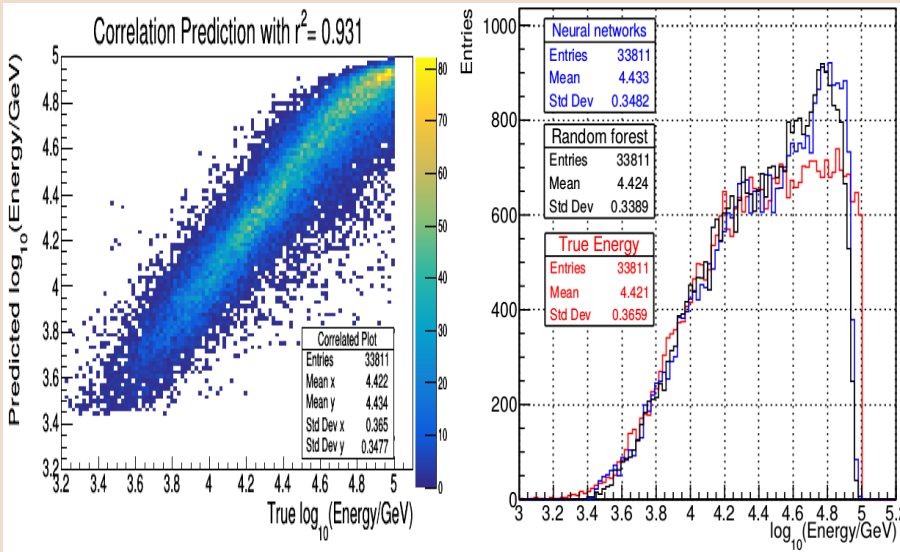
- Error absoluto medio  ≈ 0  $\approx + / - \infty$:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n} \quad (2)$$

- Bias, desviación o sesgo  ≈ 0  $\approx + / - \infty$:

$$B \equiv \langle E_{\text{Objetivo}} - E_{\text{predicho}} \rangle \quad (3)$$

¿Qué tan bien predice y qué aprendimos?



Si el tiempo lo permite...

Aprendizaje Automático en el LHC

Simulación

En el LHC se requiere simular trillones de colisiones para probar alguna hipótesis, y cada simulación puede tomar varios minutos[5].

Aprendizaje Automático en el LHC

Simulación

En el LHC se requiere simular trillones de colisiones para probar alguna hipótesis, y cada simulación puede tomar varios minutos[5].

Algunos métodos para reducir estos tiempos de simulación son:

Aprendizaje Automático en el LHC

Simulación

En el LHC se requiere simular trillones de colisiones para probar alguna hipótesis, y cada simulación puede tomar varios minutos[5].

Algunos métodos para reducir estos tiempos de simulación son:

- Usar parametrizaciones más simples.

Aprendizaje Automático en el LHC

Simulación

En el LHC se requiere simular trillones de colisiones para probar alguna hipótesis, y cada simulación puede tomar varios minutos[5].

Algunos métodos para reducir estos tiempos de simulación son:

- Usar parametrizaciones más simples.
- Usar **Redes Generativas Antagónicas (GANs)** y **Autoencoders Variacionales (VAEs)**.

Aprendizaje Automático en el LHC

Simulación

En el LHC se requiere simular trillones de colisiones para probar alguna hipótesis, y cada simulación puede tomar varios minutos[5].

Algunos métodos para reducir estos tiempos de simulación son:

- Usar parametrizaciones más simples.
- Usar **Redes Generativas Antagónicas (GANs)** y **Autoencoders Variacionales (VAEs)**.

Otra área que se puede mejorar es la **generación de eventos** [5], en donde la **optimización Bayesiana** ayuda a acelerar la generación de muestras con poco conocimiento de los detalles internos.

Aprendizaje Automático en el LHC

Disparadores

LHC: (10^9 colisiones protón-protón por segundo)[6] * (1 MB de información por colisión) = ¡1 TB de información por segundo!

¹FPGA: Field-Programmable Gate Array

²ASIC: Application Specific Integrated Circuit

Aprendizaje Automático en el LHC

Disparadores

LHC: (10^9 colisiones protón-protón por segundo)[6] * (1 MB de información por colisión) = ¡1 TB de información por segundo!

¿Cómo se enfrenta el LHC a tal cantidad de información?

¹FPGA: Field-Programmable Gate Array

²ASIC: Application Specific Integrated Circuit

Aprendizaje Automático en el LHC

Disparadores

LHC: (10^9 colisiones protón-protón por segundo)[6] * (1 MB de información por colisión) = ¡1 TB de información por segundo!

¿Cómo se enfrenta el LHC a tal cantidad de información?

El **L1T** es un filtro de eventos en tiempo real que decide en $O(1) \mu s$ si una colisión es relevante o si deberá ser descartada. Para lograr esa velocidad, el L1T hace uso de **FPGAs**¹ y **ASICs**². De las 10^9 colisiones, se obtienen $\sim 2.5 \times 10^6$ muestras relevantes [7].

¹FPGA: Field-Programmable Gate Array

²ASIC: Application Specific Integrated Circuit

¿Cómo lo hace?

Hasta hace un tiempo, se usaba un conjunto de reglas simples, lo cual reducía la precisión del disparador.

Hoy se hace uso de **Árboles de Decisión Potenciada** (BDTs), que muestran una mayor precisión respetando las restricciones de tiempo.

¿Cómo lo hace?

Hasta hace un tiempo, se usaba un conjunto de reglas simples, lo cual reducía la precisión del disparador.

Hoy se hace uso de **Árboles de Decisión Potenciada** (BDTs), que muestran una mayor precisión respetando las restricciones de tiempo.

¿Por qué FPGAs y ASICs?

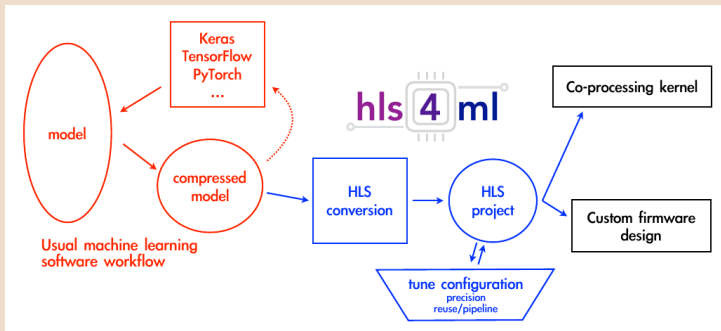
- Usan menos energía y procesan más información por segundo (Ambos).
- Son reconfigurables (FPGAs).

Desarrollo en FPGA para HEP

- Vivado HLS y Vitis AI (Xilinx).
- hls4ml[8] (CERN).

Desarrollo en FPGA para HEP

- Vivado HLS y Vitis AI (Xilinx).
- hls4ml[8] (CERN).



Aprendizaje Automático en el LHC

Reconstrucción

En una colisión no se analiza la partícula producida, sino los **productos de decaimiento** de la misma. Un mejor conocimiento de dichos productos resulta en una mayor precisión en la reconstrucción.

Anteriormente se usaban BDTs para realizar esta tarea, aunque ahora se ha volteado a mirar hacia las **Redes Neuronales Profundas** (DNNs), ya que se ajustan más a la reconstrucción; que se puede tratar como una tarea de visión computacional [5].

Aprendizaje Automático en el LHC

Asignación de incertidumbre

La **asignación de incertidumbre**[5] es un asunto crítico y juega un papel clave en la física de partículas, ya que con ella se puede corroborar la **calidad de los resultados**.

Para resolver este problema se deben unir físicos, matemáticos y computólogos y usar métodos como el muestreo Hamiltoniano Monte Carlo y métodos que puedan aprovechar el **paralelismo de las nuevas computadoras**, como la programación probabilista profunda.

An educated mind is satisfied with the degree of precision that the nature of the subject admits and does not seek exactness where only approximation is possible.

Aristotele



Oscar Roberto Chaparro Amaro



Linkedin



a170611@sagitario.cic.ipn.mx



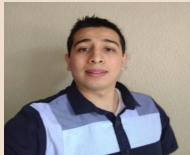
Orcid



saga9211@hotmail.com

Curriculum Vitae resumido. Perfil:

Ingeniero biomédico egresado de **UPIBI-IPN** con maestría en ciencias de la computación en **CIC-IPN** enfocado en proteómica y bioinformática con colaboración en un proyecto de rehabilitación de hueso en UPIBI, cuenta con experiencia como ingeniero de servicio en equipo médico, con colaboraciones con **FermiLab** en los proyectos de **GeantV** y **VecCore**, teniendo una estancia de investigación en el **ICN2** en nanopartículas magnéticas en aplicaciones biomédicas. Actualmente estudia un doctorado en ciencias de la computación en el CIC-IPN, colaborando con el observatorio **HAWC** en un proyecto de aprendizaje automático.



Alberto Maldonado Romo



Linkedin



amaldonador1300@alumno.ipn.mx



Orcid



alberto.maldo1312@gmaill.com

Curriculum Vitae resumido. Perfil:

Ingeniero en sistemas computacionales egresado de **ESCOM-IPN**, actualmente cursa la maestría en ciencias de la computación en **CIC-IPN** enfocado en la computación cuántica, generando un algoritmo de procesamiento de imágenes, cuenta con colaboraciones con **FermiLab** y **CERN** en los proyectos de **GeantV**. Actualmente colabora con el observatorio **HAWC** en un proyecto de aprendizaje automático.



Miguel de Jesús Martínez Felipe



Linkedin



mmartinezf2002@alumno.ipn.mx



Orcid



mjmf2402@hotmail.com

Curriculum Vitae resumido. Perfil:

Ingeniero en informática egresado de **UPIICSA-IPN** con maestría en ciencias de la computación en **CIC-IPN** enfocado al área de análisis de imágenes y máquinas de aprendizaje, cuenta con experiencia como desarrollador de cómputo en la nube en el tribunal federal de justicia administrativa, Actualmente estudia un doctorado en ciencias de la computación en el CIC-IPN, colaborando con el observatorio **HAWC** en un proyecto de aprendizaje automático.



Ricardo Balderas Paredes



rbalderas1800@alumno.ipn.mx



balderas_ricardo@hotmail.com

Perfil:

Ingeniero en Sistemas Computacionales egresado de la **ESCOM-IPN** estudiando una maestría en Ciencias e Ingeniería de Cómputo en el **CIC-IPN**, donde desarrolla un simulador de una computadora cuántica en un FPGA. Tiene experiencia como desarrollador de aplicaciones que calculan reservas para seguros y fianzas. Interesado en programación de sistemas embebidos y cómputo cuántico.

Referencias bibliográficas I



E.P. Prats, O.P. Vila, S.S. Mesquida, J.V. Marca, and National Geographic Society.

El Poder de los datos: del big data al aprendizaje profundo.
El mundo es matemático. RBA Coleccionables, 2017.



Ian H. Witten, Eibe Frank, and Mark A. Hall.

Data Mining: Practical Machine Learning Tools and Techniques.

Morgan Kaufmann Series in Data Management Systems.
Morgan Kaufmann, Amsterdam, 3 edition, 2011.



Peter Harrington.

Machine Learning in Action.

Manning Publications Co., USA, 2012.

Referencias bibliográficas II



Atulya Nagar, Durga Prasad Mohapatra, and Nabendu Chaki.
Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics: ICACNI 2015, Volume 1, volume 43.
Springer, 2015.



Kim Albertsson et al.
Machine learning in high energy physics community white paper, 2019.



LHC Collaboration.
Facts and figures about the lhc.
<https://home.cern/resources/faqs/facts-and-figures-about-lhc>.
Accessed: 2020-11-15.

Referencias bibliográficas III



Yutaro Iiyama et al.

Distance-weighted graph neural networks on fpgas for real-time particle reconstruction in high energy physics, 2020.



J. Duarte et al.

Fast inference of deep neural networks in fpgas for particle physics.

Journal of Instrumentation, 13(07):P07027–P07027, Jul 2018.