

ALGORITMOS DE CLASIFICACIÓN USANDO LA LIBRERÍA TMVA EN ROOT (TRIGGER PARA EL BOSÓN DE HIGGS)

*Juan Felipe Zapata
Jordan Hernández Daza*

*Big Data en el CERN y otros contextos
Instituto de Física
Universidad de Antioquia
2023*

Contenidos

Objetivos

- Objetivo general
- Objetivos específicos

Introducción

- ML
- Librería TMVA
- Qué es un algoritmo de clasificación

Algunos algoritmos de clasificación

- Naive Bayes
- Boosted Decision Trees
- Pykeras
- Pykeras

Implementación del Algoritmo en ROOT

Resultados

- Modelo entrenado y datos de testeo
- Comparación de los modelos

Referencias

Objetivos

Objetivo general

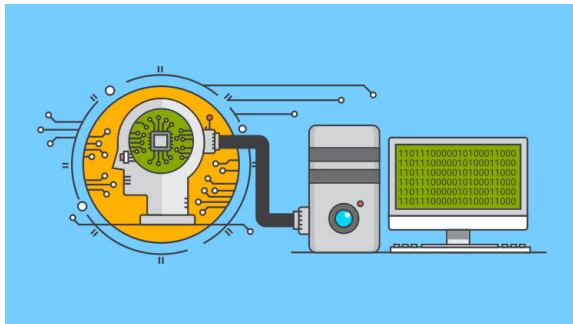
Presentar un algoritmo en ROOT (trigger) en el cual se utilizan diferentes métodos de clasificación de la librería TMVA para realizar la separación de la señal y el background en datos asociados a la masa de Bosson de Higgs.

Objetivos específicos

- ▶ Exponer los aspectos principales de cada algoritmo de clasificación.
- ▶ Importar los datos de entrenamiento y evaluación desde el repositorio "[http : //root.cern.ch/files/Higgs_data.root](http://root.cern.ch/files/Higgs_data.root)"
- ▶ Aplicar y comparar el rendimiento de los diferentes algoritmos de clasificación.

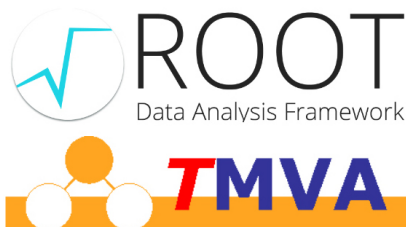
Machine learning

En computer science machine learning se define como una rama de la inteligencia artificial encargada de desarrollar algoritmos capaces de identificar patrones que se usa principalmente para clasificación y regresión.



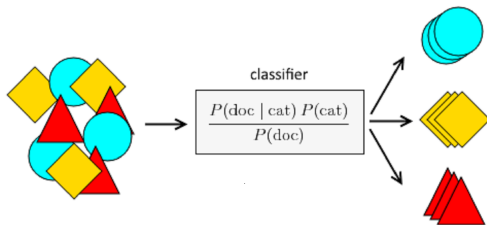
Librería TMVA

La librería TMVA (Toolkit for Multivariate Data Analysis) es una herramienta desarrollada en el entorno ROOT, análoga a Scikitlearn en Python. Proporciona una serie de técnicas y algoritmos para el análisis multivariante, incluyendo clasificación y regresión, que son ampliamente utilizados para el análisis de datos.



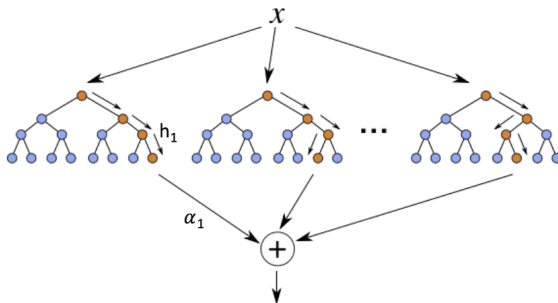
Naive Bayes

El algoritmo de Naive Bayes es un método de clasificación que se basa en el teorema de Bayes. Estima la probabilidad de que una instancia pertenezca a una clase determinada utilizando las probabilidades condicionales de los atributos. Las variables predictoras son independientes entre sí.

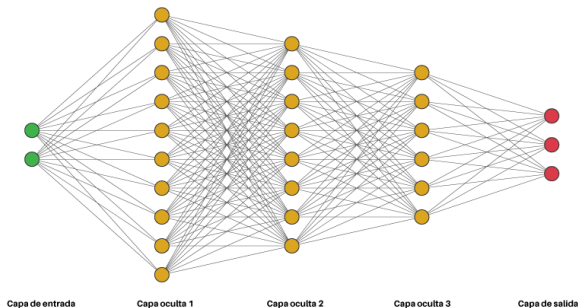


Boosted Decision Trees

Son un conjunto de modelos de árboles de decisión que se combinan y refuerzan para mejorar su capacidad de clasificación. Estos algoritmos utilizan una técnica llamada "boosting" que permite aprender secuencialmente a partir de errores anteriores. Cada árbol se ajusta a los datos de entrenamiento, poniendo más énfasis en las instancias clasificadas incorrectamente por los árboles anteriores [1].

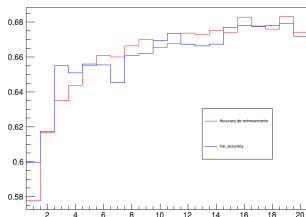
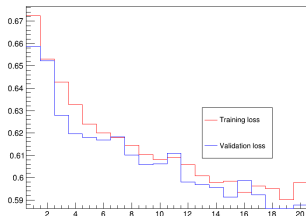


Una red neuronal es un modelo de aprendizaje automático inspirado en el funcionamiento del cerebro humano. Consiste en una colección de nodos interconectados, llamados neuronas artificiales o unidades, organizados en capas. Cada unidad toma una o más entradas, realiza un cálculo interno y produce una salida.



Métricas de la red neuronal

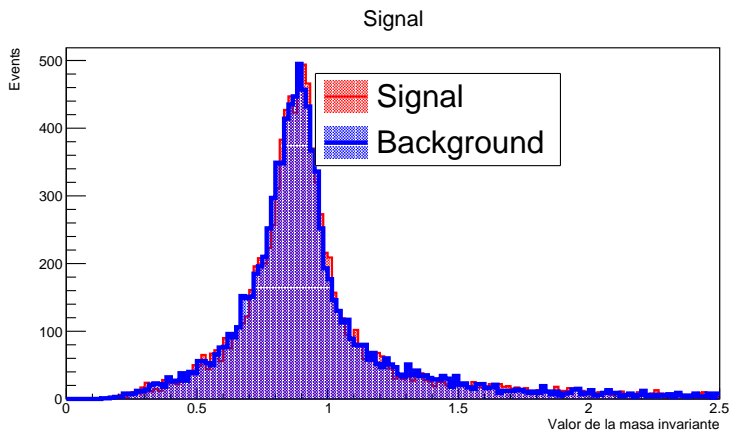
Se evalúan las métricas de pérdida y de accuracy para la red neuronal. Considerando que la función de pérdida en este caso es la entropía cruzada binaria.



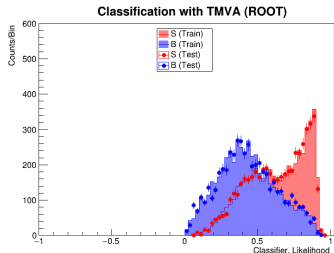
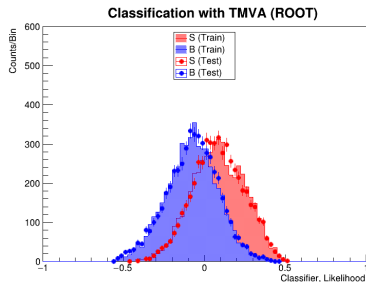
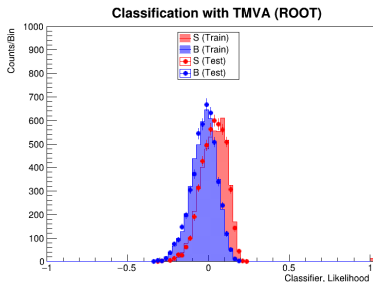
Algoritmo en ROOT [2]

Datos de entrenamiento

Tenemos 7 entradas de datos de entrenamiento



Modelo entrenado y datos de testeo



Matriz de confusión

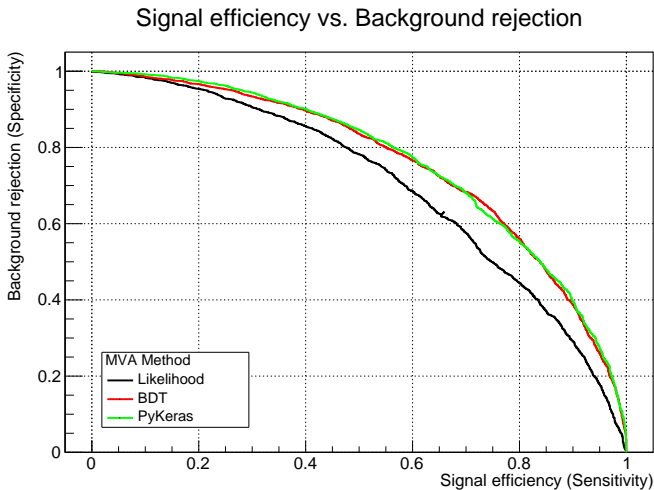
1.) La tasa de verdadero positivos (TVP, también llamada sensibilidad) se calcula como $\frac{TP}{TP+FN}$. La tasa de verdaderos positivos es la probabilidad de que un resultado positivo real dé positivo.

2.) La tasa de verdaderos negativos (también llamada especificidad), que es la probabilidad de que un resultado negativo real dé un resultado negativo. Se calcula como $\frac{TN}{TN+FP}$.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Comparación de eficiencia entre diferentes algoritmos

Verdaderos negativos vs verdaderos positivos



Referencias

[1] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189-1232. DOI: 10.1214/aos/1013203451.

[2] ROOT TMVA tutorials, [En línea]. Disponible [aquí](#). [Fecha de acceso: 6 de junio de 2023].