



1. DATOS BÁSICOS DEL TFG:

Título: Machine learning aplicado al experimento SBND

Descripción general (resumen y metodología):

El experimento SBND (Short-Baseline Near Detector) forma parte del programa de búsqueda de neutrinos estériles en Fermilab, y utiliza una cámara de proyección temporal de argón líquido (LArTPC) para detectar interacciones de neutrinos con alta precisión espaciotemporal. Una parte esencial de su misión científica es estudiar la posible aparición de neutrinos del electrón, generados a partir de neutrinos del muon a través del fenómeno de oscilación. Identificar con fiabilidad sucesos correspondientes a neutrinos del electrón entre todos los registrados resulta clave para confirmar o refutar las anomalías observadas en experimentos previos. Esta tarea implica distinguirlos de un entorno dominado por otros tipos de sucesos, como los de neutrinos del muon o de origen cósmico, cuya firma puede imitar la señal buscada.

En el ámbito del aprendizaje supervisado, una rama del machine learning, se emplea un conjunto de datos con etiquetas conocidas para entrenar un modelo capaz de realizar predicciones o tomar decisiones basadas en nuevos datos que no han sido etiquetados. Los algoritmos que se utilizan en este tipo de aprendizaje suelen apoyarse en modelos probabilísticos, como la regresión lineal, la regresión logística, los árboles de decisión o las redes neuronales, entre otros. Para seleccionar el método de clasificación más adecuado, es fundamental considerar si las variables de entrada son de tipo cuantitativo o cualitativo, así como la distribución conjunta de probabilidad de dichas variables y otras propiedades relevantes del conjunto de datos. Según estas características, se podrá optar por la técnica de clasificación que mejor se adapte al problema. Cabe señalar que el aprendizaje supervisado no es la única estrategia de aprendizaje automático utilizada en disciplinas como la física. También existen otros enfoques, como el aprendizaje no supervisado, el aprendizaje por refuerzo y diversas técnicas de aprendizaje profundo. Cada uno de estos métodos presenta ventajas y limitaciones específicas, y su idoneidad dependerá del tipo de problema que se pretenda resolver.

En este trabajo se propone aplicar técnicas de aprendizaje supervisado para desarrollar un clasificador capaz de identificar sucesos de neutrinos del electrón en los datos simulados de SBND. Se explorarán distintos algoritmos y metodologías con el fin de maximizar la eficiencia de identificación y minimizar las tasas de falso positivo, lo que repercutirá directamente en la sensibilidad del experimento.

Actividades a realizar:

- Revisión bibliográfica exhaustiva sobre técnicas para la clasificación de datos multivariantes, identificando las principales metodologías y enfoques utilizados en la literatura, con especial referencia a su estado actual, y exposición sintética del conocimiento desde una perspectiva global.
- Introducción de los conceptos teóricos necesarios, así como las herramientas a utilizar. En particular, se llevará a cabo una identificación de los aspectos conceptuales inherentes a los enfoques de clasificación objeto de estudio y su formalización matemática. Asimismo, se realizará un análisis pormenorizado de los fundamentos matemáticos que sustentan el desarrollo de la metodología objeto central del trabajo.

- Adquirir una comprensión general del funcionamiento y principios físicos de los detectores LArTPC, con énfasis en su capacidad de reconstrucción tridimensional.
- Familiarizarse con el programa experimental del SBND y su papel dentro del marco del programa Short-Baseline Neutrino.
- Analizar el contenido de los archivos de análisis de SBND y comprender la estructura de la información disponible para los sucesos.
- Identificar variables relevantes para la caracterización de sucesos de neutrinos del electrón y estudiar su poder discriminante frente a sucesos de fondo..

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

- Realizar una revisión de la literatura sobre metodologías estadísticas consolidadas para clasificación, con énfasis en su desarrollo actual y una presentación resumida del conocimiento desde una visión global en el ámbito del análisis de datos multivariantes.
- Adquirir conocimientos sobre el funcionamiento de las cámaras de deriva en argón líquido (LArTPCs) y comprender los objetivos científicos del experimento SBND, así como del programa de neutrinos de corto recorrido ("short-baseline").
- Analizar en profundidad modelos estadísticos adecuados para abordar problemas específicos, seleccionándolos en función de las características de los datos multivariantes de entrada, identificando claramente los conceptos clave, fundamentos matemáticos y metodologías asociadas.
- Comprender la estructura y contenido de los archivos utilizados en el análisis del experimento SBND.
- Llevar a cabo un estudio de viabilidad estadística para evaluar la capacidad de separación basada en las variables disponibles.
- Emplear las técnicas mencionadas anteriormente para realizar la separación de partículas en el estado final, utilizando o desarrollando herramientas computacionales y visuales según sea necesario.
- Redactar un informe final que sintetice los resultados obtenidos, las dificultades encontradas y las soluciones implementadas.

Bibliografía básica:

- C. C. Aggarwal, "Data Mining: The Textbook". Springer
- F.E. Harrel, "Regression Modeling Strategies". Springer
- B. Ratner, "Statistical and Machine-Learning Data Mining. Techniques for Better Predictive Modelling and Analysis of Big Data". CRC Press
- D. Zeltermanm "Applied Multivariate Statistics with R". Springer
- <http://sbn-nd.fnal.gov/>
- P. Machado, O. Palamara, D. Schmitz, "The Short-Baseline Neutrino Program at Fermilab", Ann.Rev.Nucl.Part.Sci. 69 (2019) 363-387. <https://www.annualreviews.org/doi/10.1146/annurev-nucl-101917-020949>
- <http://pdg.lbl.gov/2018/reviews/rpp2018-rev-neutrino-mixing.pdf>
- M. Thomson, "Modern Particle Physics". Cambridge University Press

Recomendaciones y orientaciones para el estudiante:

Se recomienda que la estudiante haya cursado las asignaturas de Estadística Multivariante y de Estadística Computacional, dado que proporcionan una base sólida para el desarrollo del trabajo. Asimismo, es deseable que la estudiante cuente con conocimientos básicos en el uso de LaTeX para la

redacción académica, y que posea autonomía en la búsqueda y selección de fuentes bibliográficas relevantes. Es imprescindible que tenga competencia en la lectura y comprensión de textos científicos en inglés, así como la capacidad para integrar información procedente de diversas fuentes con el fin de construir un discurso analítico propio y coherente.

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: JOSÉ LUIS ROMERO BÉJAR

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: jlrbejar@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos: BRUNO ZAMORANO GARCÍA

Ámbito de conocimiento/Departamento: FÍSICA TEÓRICA

Correo electrónico: bzamorano@ugr.es

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: LUNA RUBIO GOMEZ

Correo electrónico: lunarubio@correo.ugr.es