

Determinación de la influencia de las condiciones observacionales sobre el corrimiento al rojo en observaciones simuladas de DESI

Sergio David Lobo Bolaño
201218661

Director: Jaime Ernesto Forero-Romero

June 18, 2018

1 Introducción

La expansión acelerada del universo es uno de los fenómenos de mayor importancia en la astrofísica, y en general en toda la física moderna, a tal punto que impulsa gran parte de la investigación en cosmología observacional [1]. Para aumentar nuestro conocimiento en este campo, el instrumento DESI (Dark Energy Spectroscopic Instrument) se usará para conducir un mapeo durante 5 años que cubrirá un tercio del cielo [2]. El mapeo hará observaciones espectroscópicas de cuatro clases distintas de fuentes extragalácticas: muestra de galaxias brillantes (BGS), galaxias rojas luminosas (LRGs), galaxias de líneas de emisión de formación estelar (ELGs), y objetos cuasi-estelares (QSOs) [2]. El instrumento DESI es un espectrógrafo capaz de tomar hasta 5000 espectros simultáneos en un rango de 360 nm a 980 nm [3].

Las LRGs serán medidas hasta $z = 1.0$, para explorar el universo a un corrimiento al rojo incluso mayor, DESI tendrá como objetivo ELGs de hasta $z = 1.7$. Los Quásares serán observados como señales directas de la distribución de materia oscura subyacente y, a mayores corrimientos, $2.1 < z < 3.5$, por las características de absorción del bosque de Ly- α en sus espectros [2].

Antes de comenzar a funcionar, DESI necesita de simulaciones para su diseño, desarrollo y operación. Algunas de estas simulaciones se enfocan en aspectos individuales como la optimización de piezas de hardware, otras simulaciones proporcionan datos de prueba para el desarrollo de la extracción espectral antes de obtener datos reales [3]. En este proyecto en particular se usarán los resultados de las simulaciones de corrimientos al rojo de los diferentes tipos de galaxias para compararlos con los que mediría DESI desde la Tierra, y encontrar en estas simulaciones la relación esperada entre los corrimientos al rojo intrínsecos y los observados, atribuyéndole las diferencias entre ambos a las condiciones observacionales tales como el telescopio, el tiempo de exposición, luz de

la luna, etc.

La relación entre el corrimiento al rojo intrínseco de cada galaxia (simulado) y el observado (simulado) se encontrará usando algoritmos de machine learning apropiados para el problema de regresión [4], tales como regresión logística, SVRs y bosques de regresión, entre otros. Por tal razón, se escogerán tres algoritmos distintos los cuales se entrenarán, afinarán y validarán para luego escoger cuál describe mejor la relación entre ambos corrimientos, y así poder 'calibrar' el instrumento para sus futuras observaciones reales. El problema a tratar es entonces uno de aprendizaje supervisado en donde se entrenará cada algoritmo con los datos de redshift real, simulado y los parámetros mencionados anteriormente. Se espera entonces encontrar la 'función' que mapea el redshift observado al real a partir del valor de las diferentes condiciones observacionales (variables de entrada).

2 Objetivo General

Determinar la influencia de las condiciones observacionales sobre la degradación de la información de redshift en Observaciones simuladas de DESI, mediante el uso de algoritmos de machine learning

3 Objetivos Específicos

- Analizar y preprocesar los datos de entrada y salida de acuerdo al significado físico de cada variable
- Seleccionar 3 tipos de algoritmos de aprendizaje autónomo para implementar en los datos
- Entrenar y validar los algoritmos de aprendizaje supervisado con los datos de la simulación
- Comparar los resultados de los diferentes algoritmos para seleccionar el que mejores resultados obtenga

4 Metodología

Esta monografía es de carácter computacional. Los datos serán tomados de las simulaciones realizadas en el proyecto DESI y recolectados por el profesor Jaime E. Forero-Romero y se usarán recursos computacionales propios y los dispuestos por la universidad en forma física y virtual (cluster). Los algoritmos se escribirán en Python y R. Para seleccionar los algoritmos a utilizar se realizará una investigación corta que permita definir las mejores opciones capaces de ajustarse a los datos y su facilidad de implementación. El problema a resolver es de tipo regresión y aprendizaje supervisado, por

lo que el espacio de algoritmos se restringe a estas dos características. El entrenamiento y validación de los algoritmos se realizarán de acuerdo a la investigación anterior, por lo cual se deberán preprocesar los datos según lo requiera cada algoritmo. El ajuste y validación de los parámetros de cada modelo se realizará mediante validación cruzada y búsqueda exhaustiva. A partir de métricas de regresión se seleccionará finalmente el algoritmo que mejor generaliza sus resultados. El trabajo será realizado por el estudiante guiado por el profesor en dónde se revisarán los avances del proyecto semana a semana en una reunión previamente programada.

5 Consideraciones Éticas

Los datos utilizados en este proyecto serán recolectados por el Profesor Jaime E. Forero-Romero, director de este proyecto, quien hace parte de la colaboración internacional del proyecto DESI. Así mismo, los datos serán utilizados en este trabajo únicamente con los motivos de desarrollar el algoritmo anteriormente mencionado y contribuir al desarrollo del proyecto de colaboración internacional DESI. Este trabajo no debe pasar a estudio del comité de ética de la Facultad de Ciencias.

6 Cronograma

Tareas \ Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	X	X														
2	X	X	X													
3				X	X											
4				X	X	X										
5							X	X								
6							X	X	X							
7							X	X								
8										X	X					
9										X	X	X				
10													X	X	X	
11							X	X				X	X	X	X	X

- Tarea 1: Análisis preliminar y preprocesamiento de los datos
- Tarea 2: Investigación bibliográfica sobre algoritmos de machine learning adecuados
- Tarea 3: Entrenamiento Algoritmo 1

- Tarea 4: Validación Algoritmo 1 y selección de modelo
- Tarea 5: Entrenamiento Algoritmo 2
- Tarea 6: Validación Algoritmo 2 y selección de modelo
- Tarea 7: Preparación de documento y presentación 30%.
- Tarea 8: Entrenamiento Algoritmo 3
- Tarea 9: Validación Algoritmo 3 y selección de modelo
- Tarea 10: Comparación y selección de modelos
- Tarea 11: Preparación de documento y presentación

7 Personas Conocedoras del Tema

- Beatriz Eugenia Sabogal Martínez (Departamento de Física, Universidad de los Andes)
- Andres Gonzalez Mancera (Departamento de Ingeniería Mecánica, Universidad de los Andes)

References

- [1] B. D. Nord *et al.*, “SPOKES: an End-to-End Simulation Facility for Spectroscopic Cosmological Surveys,” *Astron. Comput.*, vol. 15, pp. 1–15, 2016.
- [2] A. Aghamousa *et al.*, “The DESI Experiment Part I: Science, Targeting, and Survey Design,” 2016.
- [3] A. Aghamousa *et al.*, “The DESI Experiment Part II: Instrument Design,” 2016.
- [4] C. Bishop, *Pattern Recognition and Machine Learning*. Information Science and Statistics, Springer, 2006.

Firma del Director