**Análisis estadístico (cuaderno PCA.ipynb)**

Como se mencionó antes, el análisis estadístico se realizó con datos de espectros reales medidos por el SDSS (Sloan Digital Sky Survey). En total se usaron 10 catálogos de 1000 objetos de las carpetas públicas del SDSS previamente descargados (<https://data.sdss.org/sas/dr14/eboss/spectro/redux/v5_10_0/>). Estos catálogos guardan el espectro de los objetos junto a más información que los identifica.

|  |  |
| --- | --- |
| Graphical user interface, text  Description automatically generated | Text  Description automatically generated |

De todas las observaciones se seleccionaron aquellas identificadas como cuásares y galaxias, guardándolas en arreglos por separado.

Text

Description automatically generated

Dado que cada archivo tiene diferentes rangos de medición de longitudes de onda, se estandarizó a un único rango, esto es necesario para poder hacer el PCA.

Text

Description automatically generated

Ya teniendo separados los cuásares de las galaxias, se calculó la razón entre ellas asumiendo una distribución normal de esta cantidad.

|  |  |
| --- | --- |
| Text  Description automatically generated | Chart, line chart  Description automatically generated |

En los catálogos anteriores no se almacena el corrimiento al rojo del objeto en cuestión, para obtenerlo se necesita otro catálogo de cuásares en donde sí podemos encontrar el dato, en este podemos identificar los cuásares del filtrado anterior por medio de los dos números de identificación “MJD” y   
“PLATE”.

Graphical user interface, text

Description automatically generated

Una vez filtrados, almacenamos en un arreglo la información del filtrado anterior y el corrimiento al rojo.

A screenshot of a computer

Description automatically generated with medium confidence

De todos los objetos filtrados se separaron en tres categorías para la generación de espectros de dichas tres categorías, las cuales son cuásar con z<2.1, cuásar con z>2.1 y galaxias.

Text

Description automatically generated

Para la graficación correcta de la longitud de onda de los objetos observados es necesario el corrimiento al rojo. Se almacenaron los espectros de cada categoría en arreglos separados, para sobre estos hacer el PCA.

Graphical user interface, text, application

Description automatically generated

Las cantidades totales de objetos de cada categoría son las siguientes.

Text

Description automatically generated

El PCA se realizó usando la librería sklearn a la cual se le dio como argumento los arreglos de espectros.

Text

Description automatically generatedText

Description automatically generated

Esta función regresa las componentes principales en un arreglo, las cuales graficamos con la siguiente función

|  |  |
| --- | --- |
| Text  Description automatically generated | **.**  **.**  **.** |

Para visualizar cómo las componentes principales reproducen los espectros experimentales se usó la siguiente función

|  |  |
| --- | --- |
|  |  |

Las componentes principales de cada tipo de espectro fueron guardadas en archivos de texto para usarlas después en la generación de espectros. Para generar los espectros usamos la distribución de los coeficientes de los espectros experimentales en la base de los componentes principales, estos coeficientes también fueron guardados en archivos de texto.

Text

Description automatically generated

Con estos coeficientes se generaron distribuciones de probabilidad para la generación aleatoria de dichos coeficientes para nuevos espectros, esto se hizo mediante KDE (Kernel Distribution Estimation), que lo que hace es sumar distribuciones normales colocadas en cada posición de los datos experimentales.

Chart, histogram

Description automatically generated

La imagen anterior muestra solamente las distribuciones de las primeras 9 componentes, en total se usaron las distribuciones de 50 componentes. Muestreando estas distribuciones se generan los nuevos espectros, que son vectores en la base de las primeras 50 componentes principales.

La misma idea fue aplicada para la generación del corrimiento al rojo, usando los datos de corrimiento al rojo se generó una distribución de la cual se muestrea para generar nuevos corrimientos al rojo.

Chart, histogram

Description automatically generated

**Generación de objetos (librería objast)**

Como se describió en la sección de análisis estadístico, los espectros se generaron usando los KDE obtenidos de los PCA y los coeficientes de los datos experimentales en la base de los primeros 50 PCA.

En esta librería se cargan los archivos de texto generados por el análisis estadístico y construye las distribuciones de las que muestrearemos para generar espectros y corrimientos al rojo.

Text

Description automatically generated

Text

Description automatically generated

Para cada tipo de objeto que generaremos se creó una clase, pero antes se definió una clase padre “objast” (objeto astronómico) que tiene como atributo su posición en coordenadas (RA, DEC) (ascensión recta y declinación), también tiene como método la generación de posiciones aleatorias homogéneas sobre la cúpula celeste.

Text

Description automatically generated

Después, se definen las clases “oamas” y “oamenos” (objeto astronómico con z>2.1 y z<2.1 respectivamente), estas clases heredan de la clase objast y tienen como atributo propio el corrimiento al rojo “z” para el cual tienen un método que lo genera de acuerdo con la distribución de corrimiento al rojo que se obtuvo del análisis estadístico.

|  |  |
| --- | --- |
| Text  Description automatically generated |  |

Y finalmente se declaran las clases los tipos de objetos que se generarán en el programa final que son:

* Cuásar con z<2.1
* Cuásar con z>2.1
* Galaxia con z<2.1
* Galaxia con z>2.1

Que tienen como atributos su espectro, tipo de objeto y tipo de espectro, aparte de tener el método que genera su espectro en función de su atributo tipo de espectro. Notar que la generación de espectros de galaxias es la misma para z<2.1 y z>2.1, la única diferencia entre estas dos clases es la herencia del método de generación de corrimiento al rojo, uno hereda de “oamas” y otro de “oamenos”.

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated