**Contexto del problema**

Tengo un conjunto de curvas de luz de diferentes tipos de estrellas variables:

* Cepheid,
* RR Lyrae (RRL),
* Estrellas Binarias (EB).

Estas estrellas variables muestran cambios periódicos en su brillo (representado por el FLUXCAL), que se mide a lo largo del tiempo (MJD). El objetivo es clasificar correctamente entre los diferentes tipos de estrellas utilizando sus curvas de luz.

Una curva de luz es una serie temporal de mediciones de flujo en diferentes puntos de tiempo. Las estrellas variables tienen diferentes patrones de periodicidad en sus curvas de luz, lo que las distingue.

Para poder clasificar correctamente estas curvas de luz, se necesita extraer *características clave* de las curvas que el modelo de clasificación pueda usar. Una de las características más importantes en una curva de luz es su período (es decir, el tiempo que tarda el patrón de brillo en repetirse). Aquí es donde entra el Lomb-Scargle Periodogram.

**Qué es el Lomb-Scargle Periodogram y por qué es importante**

El Lomb-Scargle Periodogram es una técnica usada para detectar periocidad en datos que son irregularmente muestreados (es decir, cuando los datos no están igualmente espaciados en el tiempo, como en muchas observaciones astronómicas). Este es un método especializado que extiende el análisis de la transformada de Fourier para manejar datos desigualmente espaciados, que es una característica común en las curvas de luz astronómicas.

El período de una curva de luz es la característica que te dice cada cuánto tiempo la estrella completa un ciclo completo de variación en su brillo. Este período es diferente para distintos tipos de estrellas, y es un rasgo clave que puedes usar para clasificarlas.

Ejemplo de períodos típicos:

* Cepheid: Tienen períodos de días a semanas.
* RR Lyrae: Tienen períodos mucho más cortos, típicamente alrededor de 0.5 días.
* Estrellas Binarias: Pueden tener períodos variables dependiendo de la separación de las estrellas binarias.

**¿Qué datos se tienen hasta el momento?**

El pipeline procesa datos fotométricos que consisten en:

* MJD: Tiempo en días julianos modificados. Esto representa los puntos de tiempo en los que se tomó la medición.
* FLUXCAL: El flujo calibrado de la estrella en ese tiempo, que mide el brillo.
* BAND: El filtro fotométrico utilizado (se esta trabajando con la banda i, que representa la luz infrarroja cercana).
* label: La etiqueta de clasificación que indica si la curva de luz pertenece a un Cepheid (0), una RR Lyrae (1), o una Estrella Binaria (2).

El objetivo es clasificar las estrellas según su tipo, y el período de la curva de luz es una característica crucial para esta clasificación.

**Cómo encaja el Lomb-Scargle Periodogram en el pipeline**

El Lomb-Scargle Periodogram ayudará a calcular el mejor período para cada curva de luz, que luego se puede usar como una característica adicional al entrenar el modelo de clasificación.

**Aplicación del Lomb-Scargle Periodogram**

¿Qué hace el Lomb-Scargle Periodogram?

1. Entrada: Toma los datos de la curva de luz —es decir, una secuencia de valores de tiempo (MJD) y brillo (FLUXCAL)—.
2. Cálculo del mejor período: A través de la técnica de Lomb-Scargle, encuentra el período dominante que mejor ajusta una sinusoide a los datos.
3. Salida: El mejor período, que se puede usar como una característica adicional para clasificar la curva de luz.

-> ¿Por qué usarlo en este problema? El período es una característica distintiva de los diferentes tipos de estrellas variables. Por ejemplo, las RR Lyrae tienen períodos de alrededor de 0.5 días, mientras que los Cepheids pueden tener períodos mucho más largos. El período puede ser una señal clave que distinga una clase de estrella de otra. Al agregar el período como una característica al conjunto de datos, se está proporcionando al modelo información adicional que puede mejorar su capacidad de clasificación.

Fuente: <https://www.astroml.org/gatspy/periodic/lomb_scargle.html>