

Mi portafolio y un ejemplo en series de tiempo

Dra. Darlin Soto

Universidad de la Frontera

03/10/2022

¿Qué veremos hoy?

- 1 Portafolio
 - Motivación
 - ¿Qué es y qué contiene un portafolio?
 - ¿Por qué es importante tener uno?
 - Tipos de proyectos en un portafolio
 - ¿Dónde crear un portafolio?
 - Ejemplo

- 2 Proyecto
 - Aspectos importantes en un proyecto
 - Ejemplo de proyecto: Pronosticando el número de manchas solares
 - Conceptos en series de tiempo

Ejemplos de portafolios (en modelaje)



Referencias:

<https://www.pinterest.ca/pin/636414991076374423/>

<https://www.pinterest.ca/pin/441212094759672608/>

¿Qué es y qué contiene un portafolio?

Un portafolio se define como una **evidencia pública de tus habilidades creativas, técnicas y blandas, su enfoque para analizar datos de manera efectiva y obtener información, y su capacidad para comunicar el resultado a la audiencia**. Un portafolio puede contener, entre otras cosas, la construcción de un algoritmo desde cero, investigación de nuevos métodos, análisis significativos, etc., todas creadas por ti.

¿Por qué es importante tener un portafolio?

Existen dos razones principales:

- 1 Es una **herramienta clave en el proceso de contratación**, ya que permite mostrar tus habilidades y experiencia en proyectos anteriores.
- 2 Es una gran **oportunidad de aprendizaje** pues puedes llevar a cabo proyectos personales que satisfagan tu curiosidad por aprender cosas nuevas.

Tipos de proyectos en un portafolio

Existen dos tipos de proyectos:

- ➊ **Proyectos basados en código:** Replica proyectos de ciencia de datos del mundo real tomando un conjunto de datos y resolviendo un problema a su alrededor.
Ejemplo: Extraer un conjunto de datos y realizar algún análisis o entrenar un modelo.
- ➋ **Proyectos basados en contenido:** Basados en explicar algún contenido, son extremadamente efectivos para mostrar sus habilidades de comunicación y escritura.
Ejemplo: Tutoriales que expliquen algún concepto para audiencia no técnica o que muestren como funciona una herramienta en particular.

¿Dónde crear un portafolio?

Existen varios sitios en internet gratuitos donde puedes crear tu portafolio, como por ejemplo:

- GitHub
- Deepnote
- DAGsHub
- Kaggle

Nos enfocaremos en **GitHub** por su popularidad y variedad de cosas que puedes hacer. GitHub es de carácter colaborativo global donde las personas comparten y colaboran en proyectos. Además, le permite a la gente ver lo que has construido y cómo lo has construido.

Darlin Soto



[LinkedIn](#) | [GitHub](#)

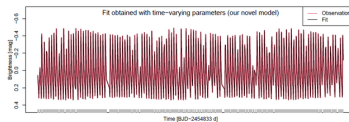
I am Doctor in Statistics working at Universidad de la Frontera, Chile.

I am developing statistical models for astronomical objects.

Portfolio

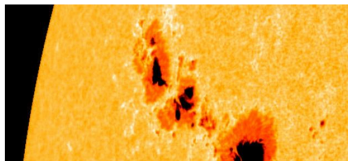
Model for Periodic Variable Stars

This repository includes codes in R software to reproduce the figures in the paper Motta, G., **D. Soto**, and M. Catelan (2022, January). Periodic variable stars modulated by time-varying parameters. ApJ 925(1), 73.



Pronosticando el número de manchas solares

Este repositorio incluye el ajuste de un modelo ARMA para pronosticar el número de manchas solares.



https://darlinsoto.github.io/Darlin_Portfolio/

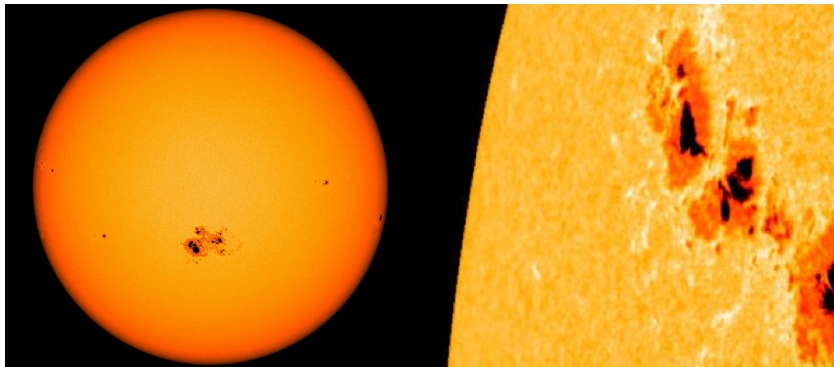
7 Aspectos importantes en un proyecto

1. **Sé auténtico y persigue tu pasión:** Los mejores proyectos no son los que utilizan las herramientas y modelos más recientes o más complejos, sino los que provienen de un lugar de auténtica pasión.
2. **La idea:** ¿Qué problema quiero resolver?, ¿quién se beneficiaría de mis análisis y/o producto?, etc.
3. **Cuenta una historia:** Asegurarte de que los lectores puedan seguir tu trabajo de principio a fin. Proyecto conciso pero cautivador.
4. **Muestra tus habilidades técnicas, pero evita la sobreinformación:** Un buen proyecto demuestra tus habilidades técnicas, pero eso no significa que debas aplicar todas las habilidades técnicas que sabes.

7 Aspectos importantes en la creación de un portafolio

5. **Fuente de datos:** Evita los conjuntos de datos de juguete como Titanic, MNIST o Iris. Descarga desde internet conjuntos de datos que sean interesantes para ti.
6. **No descuides tus habilidades interpersonales:** Explicar un problema complejo de manera simple y concisa.
7. **Diseña para tus lectores:** Guías a los lectores a la información relevante sin abrumarlos, al mismo tiempo que les brindas la oportunidad de profundizar más si así lo desean.

Ejemplo de proyecto: Pronosticando el número de manchas solares



Referencia: <https://en.wikipedia.org/wiki/Sunspot>

Datos obtenidos del libro Brockwell, Peter & Davis, Richard. (2002). An Introduction to Time Series and Forecasting (ejemplo 3.2.9).

Desarrollaremos nuestro proyecto en Jupyter notebook! Pero antes... Necesitamos saber algunos conceptos de series de tiempo.

Conceptos en series de tiempo

Serie de tiempo: Es una sucesión de datos medidos en determinados momentos y ordenados cronológicamente,

$$\{(t = 1, X_1), (t = 2, X_2), (t = 3, X_3), \dots, (t = N, X_N)\}.$$

Ejemplo de series de tiempo:

- La temperatura en un observatorio meteorológico medida sucesivamente cada día

$$\{(t = \text{día } 1, \text{Temp}_1), (t = \text{día } 2, \text{Temp}_2), \dots, (t = \text{día } N, \text{Temp}_N)\}.$$

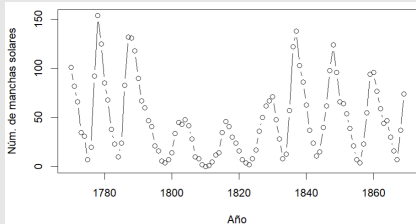
- El peso de una persona medido sucesivamente en el consultorio médico

$$\{(t = \text{Cita } 1, \text{Peso}_1), (t = \text{Cita } 2, \text{Peso}_2), \dots, (t = \text{Cita } N, \text{Peso}_N)\}.$$

Conceptos en series de tiempo

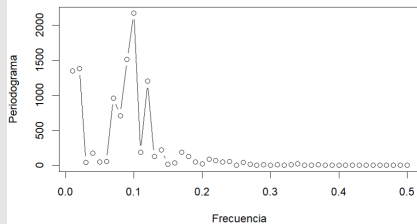
Las series de tiempo se pueden estudiar en dos dominios:

Dominio del tiempo



Ejemplo: modelos ARMA

Dominio de la frecuencia



Ejemplo: Análisis de Fourier

El periodograma evaluado en la frecuencia f_n se define en función de la Transformada de Fourier Discreta (DTF) como:

$$I_y(f_n) = \frac{1}{2\pi N} \left| \sum_{t=1}^N y_t \exp(i2\pi f_n t) \right|^2, \quad f_n = n/N, \quad n = 1, \dots, N/2.$$

Modelo ARMA(p, q)

Un proceso ARMA $\{y_t\}$ con media 0 y orden p, q es definido como solución de la ecuación

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q},$$

donde ϵ_t es un ruido blanco con media 0 y varianza σ_ϵ^2 . Los parámetros del modelo son ϕ, θ , y σ_ϵ^2 .

Para que el proceso sea estacionario, es necesario que las raíces r_1, \dots, r_p del polinomio $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ cumplan $|r_j| > 1$.

La densidad espectral (PSD) de un proceso ARMA(p, q) es dada por:

$$P_y(f) = \sigma_\epsilon^2 \frac{\left| 1 + \sum_{j=1}^q \theta_j \exp([2\pi if]^j) \right|^2}{\left| 1 - \sum_{k=1}^p \phi_k \exp([2\pi if]^k) \right|^2}, \quad 0 \leq f \leq 0.5.$$

Relación entre el periodograma y los procesos ARMA

El periodograma es un estimador de la densidad espectral de los procesos ARMA, esto es,

$$E[I_y(\omega)] \rightarrow P_y(f).$$

Ahora estamos listos para ir a realizar nuestro proyecto!

Referencias

- <https://www.kdnuggets.com/2021/10/strong-data-science-portfolio-as-beginner.html>
- <https://towardsdatascience.com/creating-a-data-science-portfolio-bd485382f49>
- <https://towardsdatascience.com/how-to-build-a-data-science-portfolio-5f566517c79c>
- <https://www.analyticsvidhya.com/blog/2021/04/how-to-make-an-impressive-data-science-portfolio/>
- <https://analyticsindiamag.com/how-to-build-a-data-science-portfolio-in-college/>
- <https://www.datacamp.com/blog/how-to-build-a-great-data-science-portfolio-with-examples>
<https://towardsdatascience.com/modelling-the-number-of-sunspots-with-time-series-analysis-39ce7d88cff3>