



TAREA 10

Fecha de entrega: 6/12/2018 23:59 hrs

Problema 1

El archivo adjunto `correlacion.dat` contiene 3 columnas:

- `year`: el año de la medición.
- `N_MPhD`: el número total de Ph.D.'s en Matemáticas otorgados durante el correspondiente año en Estados Unidos.
- `U`: la reserva de Uranio enriquecido (peso en libras) almacenado en plantas nucleares de Estados Unidos durante el correspondiente año (no pregunten cómo lo averigué, pero los datos son reales).

Queremos responder 2 preguntas sutilmente distintas:

- 1) ¿Cuánto deben bajar las reservas de Uranio para que nadie obtenga un doctorado en Matemáticas?
- 2) ¿Cuántos Ph.D. se deberían graduar en un año para que las reservas de Uranio bajaran a cero?

Para responder estas preguntas, necesitamos extrapolar la correlación que existe entre estas dos cantidades.

Responda las siguientes preguntas:

- (a) El archivo contiene 13 filas, es decir, 13 puntos. ¿Por qué es mala idea fitear un polinomio de orden 12 a los puntos para obtener respuesta a nuestras preguntas 1) y 2)?
- (b) Responda la pregunta 1) ajustando un polinomio $p = p(U)$ de orden 1 a los 13 puntos y encontrando dónde es que el polinomio intersecta al eje y ($=N_MPhD$). (Puede que la respuesta sea una cantidad de Uranio negativa, es decir, imposible que no haya doctorados en matemáticas). Para hacer este ajuste defina una función de mérito como χ^2 y busque su mínimo. Tiene libertad de usar el método que quiera pero describa en detalle qué hizo.
- (c) Responda la pregunta 2) ajustando un polinomio $p = p(N_MPhD)$ de orden 1 a los 13 puntos y encontrando donde es que el polinomio intersecta al eje y ($=U$). Mismo método que en la parte anterior pero ajustando U como función de N_MPhD en vez de N_MPhD como función de U .
- (d) Ahora dese cuenta de que los polinomios de las partes b y c no son equivalentes. Intente responder la pregunta 1 usando el polinomio de la parte c y la pregunta 2 usando el polinomio de la parte b y verá que no son equivalentes (también puede graficar los dos polinomios juntos). ¿Por qué es que no son equivalentes?
(Si quiere, puede intentar escribir una función de mérito tipo χ^2 pero que sí sea simétrica, es decir, que el resultado sea independiente de si ajustamos N_MPhD como función de U o si ajustamos U como función de N_MPhD .)
- (e) Explore qué pasa cuando usa polinomios de orden mayor para responder las mismas preguntas y explique sus resultados. Explore al menos un polinomio de orden 5 y estudie su comportamiento.

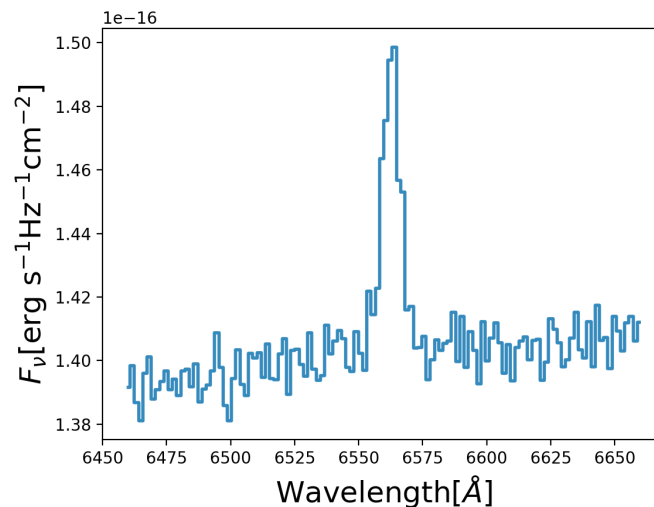
***IMPORTANTE:** Obviamente nada de esto tiene sentido, correlación no implica causalidad.

Problema 2

En el problema anterior ajustamos un polinomio a un set de datos, un problema clásico de regresión lineal que puede resolverse de manera exacta invirtiendo una matriz. En el problema a continuación exploramos una función no-lineal. En este caso necesitamos otro algoritmo distinto para la minimización, por ejemplo, un algoritmo tipo Levenberg-Marquardt.

La técnica de la espectroscopía, que se usa en múltiples disciplinas, consiste en estudiar la radiación emitida por una fuente como función de la longitud de onda. Las características de los espectros observados tales como la intensidad y forma del continuo y de las líneas de emisión y absorción, nos permiten entender las propiedades físicas del ente emisor de la radiación.

En la figura a continuación, se observa un segmento del espectro de una fuente que muestra radiación continua (con una leve pendiente) y una línea de emisión. Las unidades son de energía emitida por unidad de tiempo por unidad de área por unidad de frecuencia f_ν [erg s⁻¹Hz⁻¹cm⁻²] vs. longitud de onda en [Angstrom]. Su trabajo consiste en modelar simultáneamente el continuo y la línea de absorción de la figura (los datos los encontrará en el archivo **espectro.dat**).



Las líneas de emisión son en teoría casi infinitamente delgadas (hay un ensanchamiento intrínseco dado por el principio de incertidumbre pero es muy pequeño). Las observaciones, sin embargo, siempre muestran líneas mucho más anchas. Dependiendo del mecanismo físico que produce el ensanchamiento, la forma de la línea será distinta. Ud. deberá modelar la línea asumiendo el mecanismo de ensanchamiento más típico, el cual produce líneas gaussianas.

El modelo completo será el de una línea recta para el continuo (2 parámetros) más una función gaussiana con 3 parámetros: amplitud, centro y varianza. Es decir, debe modelar 5 parámetros a la vez.

`scipy.stats.norm` implementa la función Gaussiana si no quiere escribirla Ud. mismo. La forma de usarla es la siguiente: `g = A * scipy.stats.norm(loc=mu, scale=sigma).pdf(x)`; donde `x` es la longitud de onda donde evaluar la función.

Produzca un gráfico que muestre el espectro observado y el mejor fit obtenido. Provea una tabla con los mejores parámetros, su error estándar y el valor de χ^2_{reducido} en su mínimo.

Para buscar el mejor modelo recuerde que es importante dar un punto de partida que sea cercano para que los algoritmos converjan de manera efectiva. Ud. debe idear un método para buscar ese punto de partida.

Problema 3

En la Tarea nro. 3. utilizamos datos del *Goddard Institute for Space Science* (GISS) sobre la temperatura del planeta. Más específicamente, utilizamos el archivo `GLB.Ts+dSST-short.csv` que contiene información sobre las anomalías de temperatura medidas en la Tierra y los océanos (vea la Tarea 3 y la página oficial <https://data.giss.nasa.gov/gistemp/> para más detalles).

Utilizando estos mismos datos (versión más completa `GLB.Ts+dSST.csv`), y asumiendo que no hacemos nada para alterar la tendencia mostrada por los datos, estime para qué año la temperatura habrá cambiado en 2 grados celsius.

Explique en detalle cómo hizo la estimación (por ejemplo, ¿qué forma paramétrica escogió?) y qué algoritmo de minimización usó. Provea un intervalo de confianza para el año de la catástrofe más grande en la historia de la humanidad ([artículo random al respecto](#)).

Instrucciones importantes.

- **Por 0.5 ptos. extra** utilizables en esta tarea. Hay una imagen escondida en el repositorio (en un commit antiguo). Revise la bitácora (`git log`) del repositorio para averiguar donde está. Recupérela sin usar github web (averigue cómo hacerlo, hay multiples tutoriales en la web) y agréguela al final de su informe. Describa los pasos que siguió para recuperar la imagen. Recuerde que no puede simplemente buscarla en la interfaz web de github, la idea es que aprenda a recuperar archivos en un repositorio cualquiera de git (independiente de github.com).
- Repartición de puntaje: 40 % implementación y resolución de los problemas (independiente de la calidad de su código); 45 % calidad del reporte entregado: demuestra comprensión de los problemas y su solución, claridad del lenguaje, calidad de las figuras utilizadas; 5 % aprueba a no PEP8; 10 % diseño del código: modularidad, uso efectivo de nombres de variables y funciones, docstrings, uso de git, etc. Los 3 problemas tienen igual peso en la nota final.
- El informe debe ser entregado en formato `pdf`, este debe ser claro sin información de más ni de menos. **Esto es muy importante, no escriba de más, esto no mejorará su nota sino que al contrario.** La presente tarea probablemente no requiere informes de más de 5 páginas. Asegúrese de utilizar figuras efectivas y tablas para resumir sus resultados. Revise su ortografía.
- Evaluaremos su uso correcto de `python`. Si define una función relativametine larga o con muchos parámetros, recuerde escribir el *docstring* que describa los parámetros que recibe la función, el output, y el detalle de qué es lo que hace la función. Recuerde que generalmente es mejor usar varias funciones cortas (que hagan una sola cosa bien) que una muy larga (que lo haga todo). Utilice nombres explicativos tanto para las funciones como para las variables

de su código. El mejor nombre es aquel que permite entender qué hace la función sin tener que leer su implementación ni su *docstring*.

- Su código debe aprobar la guía sintáctica de estilo (PEP8). En [esta página](#) puede chequear si su código aprueba PEP8.
- Utilice `git` durante el desarrollo de la tarea para mantener un historial de los cambios realizados. La siguiente [cheat sheet](#) le puede ser útil. **Revisaremos el uso apropiado de la herramienta y asignaremos una fracción del puntaje a este ítem.** Realice cambios pequeños y guarde su progreso (a través de *commits*) regularmente. No guarde código que no corre o compila (si lo hace por algún motivo deje un mensaje claro que lo indique). Escriba mensajes claros que permitan hacerse una idea de lo que se agregó y/o cambió de un `commit` al siguiente.
- Al hacer el informe usted debe decidir qué es interesante y agregar las figuras correspondientes. No olvide anotar los ejes, las unidades e incluir una *caption* o título que describa el contenido de cada figura.
- La tarea se entrega subiendo su trabajo a github. Trabaje en el código y en el informe, haga *commits* regulares y cuando haya terminado asegúrese de hacer un último `commit` y luego un `push` para subir todo su trabajo a github. **REVISE SU REPOSITORIO PARA ASEGURARSE QUE SUBIÓ LA TAREA. SI UD. NO PUEDE VER SU INFORME EN GITHUB.COM, TAMPOCO PODREMOS NOSOTROS.**