

Interpolación: problemas

1. Descarga desde la web de la asignatura (estadistica-dma.ulpgc.es/FCC) en la sección “Descargas” el archivo `sstCanariasMissing.csv`. Este archivo contiene dos variables, fecha y temperatura media de la superficie del mar en la zona geográfica del archipiélago canario. Puedes comprobar (utilizando excel, por ejemplo), que hay muchas fechas en las que no se dispone del registro de temperatura. El objetivo de este primer problema es completar dicho registro rellenando las temperaturas faltantes mediante interpolación.
 - (a) Abre este archivo con matlab (usa la función `readtable`) y construye un gráfico representando la temperatura (eje Y) frente a la fecha (eje X).
 - (b) Determina cuántos valores perdidos contiene la variable temperatura.
 - (c) Utiliza la función `interp1()` de matlab para interpolar los valores de temperatura faltantes mediante interpolación lineal. Ten en cuenta que la llamada a esta función es de la forma `interp1(x,y,xq)`, donde `x` e `y` son las observaciones completas que se van a utilizar como base para la interpolación, y `xq` es el vector de valores de `x` (en este caso fechas) donde se quieren calcular los valores interpolados de `y` (en este caso temperatura media del agua de mar)
 - (d) Repite el apartado anterior pero utilizando ahora interpolación spline.
 - (e) Representa en el mismo gráfico los valores originales y los interpolados linealmente (para visualizar como se han rellenado los huecos)
 - (f) Idem, pero con los valores interpolados mediante splines.
 - (g) Descarga ahora el archivo `sstCanarias.csv`. Este era el archivo original con las temperaturas media del agua de mar en Canarias, antes de que se perdieran datos. Por tanto aquí las observaciones están completas. Para evaluar cuanto se parecen los valores interpolados a los valores realmente observados, construye la función `compara(obs, interp)` que reciba como argumentos los valores observados originales y los obtenidos mediante interpolación y que devuelva la distancia (euclídea) que los separa. Esto es, si los valores observados son $O = (o_1, o_2, \dots, o_n)$ y los interpolados son $I = (i_1, i_2, \dots, i_n)$ la función debe devolver:

$$\text{compara}(O, I) = \sqrt{\sum_{k=1}^n (o_k - i_k)^2}$$

Calcula la distancia entre los observados y los reconstruidos por interpolación lineal, así como la distancia entre los observados y los reconstruidos mediante splines. ¿Cuál de las dos interpolaciones ha producido resultados más próximos a los realmente observados?

- (h) Calcula la temperatura media de agua de mar en los años 2009, 2010 y 2011 (el año 2012 no lo consideramos porque solo hay datos hasta junio) usando la base de datos que contiene valores perdidos (ten en cuenta que la función `'year()'` permite extraer el año de una variable fecha). Repite la misma tarea con la base de datos que tiene los datos completos. Cuenta cuántos valores perdidos hay cada año. ¿Crees que la diferencia entre las medias obtenidas con y sin valores perdidos depende del número de valores perdidos que hay cada año?
- (i) Calcula y representa gráficamente la temperatura media mensual durante el periodo observado. ¿Cuales son los meses en los que se observan, respectivamente, los valores promedio mínimo y máximo de temperatura superficial del agua en Canarias?