

Tarea 1 - Data Analysis

October 19, 2018

1 Introducción

Esta tarea tiene como tema el análisis de datos, que es el proceso de inspeccionar, limpiar y transformar datos con el objeto de resaltar información útil. Vamos a enfocar la tarea en el análisis de series de tiempo, por lo que trabajaremos con un dataset famoso que registra la concentración atmosférica de co2 en Mauna Loa entre los años 1958 y 2001.

2 Tarea

1. Descargue el dataset mauna-loa-atmospheric-co2 que está disponible en la librería `statmodels` y accediendo como `statsmodels.api.datasets.co2.load_pandas().data`
2. Diseñe un estilo por defecto para los gráficos que al menos cambie colores, tamaño imagen y tamaño textos. Grafique la serie temporal y el histograma de concentraciones.
3. Agregue la columna llamada 'days' que registre la cantidad de días desde el primer registro.
4. Interpolar los valores NaN usando un estimador spline en la cantidad de días. Grafique.
5. Calcule y grafique un estimador lineal y un estimador cuadrático para la serie. Compare los errores de ajuste.
6. Con los residuos del modelo de menor error, realizar un análisis de frecuencia para encontrar el (los) periodo(s) más importante(s). Grafique.
7. Seleccionando el periodo más importante p , ajuste un modelo de parámetros a, b, c y d de la forma $m(t) = a + b \sin(\frac{2\pi ct}{p} + d)$ a los datos utilizando mínimos cuadrados. Grafique.
8. Con los residuos de este modelo, grafique la serie y su histograma, y calcule errores. Ajuste una distribución normal a los residuos, y repita con una distribución no simétrica. Grafique.
9. Construya un DataFrame con todos los errores de los modelos. Guardelo en formato hdf5 y luego carguelo desde el disco. Grafique.
10. Construya un gráfico final de su análisis y guardelo en formato pdf.

3 Entrega

El formato de entrega será subir a u-cursos el informe `tarea1.pdf` y el Jupyter notebook `tarea1.ipynb`, que debe ejecutar sin errores desde la primera celda a la última, donde cada celda debe resolver cada uno de los pasos descritos. Todo el código debe estar en el mismo notebook, el código debe estar comentado y el notebook debe estar escrito en forma de informe técnico, escribiendo una celda markdown antes de cada celda de código. Una vez que su notebook ejecute correctamente, deberá descargar la versión pdf del notebook y adjuntarla a su tarea. Este archivo corresponde a su informe de tarea, que deberá llamarse `tarea1.pdf`.