



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada



SERIES TEMPORALES Y MINERÍA DE FLUJOS DE DATOS

E.T.S. de Ingenierías Informática y de
Telecomunicación

Trabajo Autónomo I

Series Temporales

**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN E
INTELIGENCIA ARTIFICIAL
UNIVERSIDAD DE GRANADA**



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada



1. Introducción

Se presentan al alumno 826 ficheros de datos (carpeta **“datos/DatosEstaciones - 2018-02”**), correspondientes a datos meteorológicos medidos por estaciones de la AEMET (Agencia Estatal de Meteorología) repartidas por toda España, desde Mayo de 2013 hasta Febrero de 2018. Cada fichero tiene el nombre XXXXX.csv, donde XXXXX representa el código de la estación meteorológica que proporciona los datos. Para saber el nombre de la estación concreta asociada a cada archivo, puede consultarse el archivo Excel **“ListadoEstaciones2017-11.xlsx”** de la carpeta **“datos”**.

El contenido de cada fichero asociado a una estación meteorológica es texto plano. Cada línea contiene la siguiente información (ver fichero **“Columnas-Unidades.txt”** de la carpeta **“datos”**):

- Columna 1 : Identificador Estación
- Columna 2 : Fecha
- Columna 3 : Temperatura Máxima (°C)
- Columna 4 : Hora Temperatura Máxima
- Columna 5 : Temperatura mínima (°C)
- Columna 6 : Hora Temperatura mínima
- Columna 7 : Temperatura Media (°C)
- Columna 8 : Racha máxima de viento (Km/h)
- Columna 9 : Hora de Racha Máxima
- Columna 10 : Velocidad media de Viento (Km/h)
- Columna 11 : Hora de Velocidad Máxima de viento
- Columna 12 : Precipitación Total diaria (mm)
- Columna 13 : Precipitación de 0 a 6 horas (mm)
- Columna 14 : Precipitación de 6 a 12 horas (mm)
- Columna 15 : Precipitación de 12 a 18 horas (mm)
- Columna 16 : Precipitación de 18 a 24 horas (mm)

Cada columna en cada fila del fichero está separada por caracteres punto y coma ‘;’. Los datos se proporcionan en crudo y al completo, tal y como se obtienen directamente desde la web de la AEMET <https://datosclima.es/Aemet2013/DescargaDatos.html> . Se corresponden con datos diarios de temperaturas mínima y máxima, rachas de viento máxima, velocidad media del viento diaria, y precipitaciones captadas por la estación (total diaria o separada por franjas horarias). No obstante, debido a problemas de medición en los sensores, fallos eléctricos, cortes por mantenimiento, etc., es posible que alguna o algunas estaciones contengan valores perdidos en algún día puntual o en ráfagas de días.

El objetivo de la práctica de Series Temporales consiste en que el alumno, de forma individual, seleccione una estación meteorológica concreta, y trate de predecir los valores siguientes de la serie de temperatura máxima, tanto a escala diaria como a escala mensual.



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada



2. Descripción del trabajo a realizar

El trabajo a realizar es individual, no permitiéndose realizarlo por parejas o grupos. Se pide:

- o Seleccionar una estación meteorológica concreta, según el fichero Excel comentado en el apartado anterior.
- o Responder a la siguientes preguntas:
 - ¿Qué valores de temperatura máxima, a escala mensual, se espera que tengan los meses de Marzo y de Abril de 2018?
 - ¿Qué valores de temperatura máxima, a escala diaria, se espera para la primera semana de Marzo de 2018?

Nótese que, a partir de los datos de la estación meteorológica, el alumno deberá crear dos series distintas de temperaturas máximas: Una con valores de temperatura a escala mensual (un único valor de temperatura por mes), y otra a escala diaria (un único valor de temperatura por día).

El alumno deberá seguir una metodología de análisis y predicción de series temporales utilizando las técnicas estudiadas durante las sesiones presenciales de teoría y prácticas, así como hacer uso de otras técnicas de análisis de datos (preprocesamiento, agregación, tratamiento de valores perdidos, etc.) que el alumno estime oportunas, tanto para rellenar valores perdidos como para agregar los datos a escala mensual. Deberá, a su vez, justificar detalladamente cada uno de los pasos y decisiones a tomar durante el proceso de modelado y predicción. En concreto, **se deberán abordar los siguientes ítems:**

- Preprocesamiento (por ejemplo: tratamiento de valores perdidos, agregación de datos para la serie de escala mensual, selección de datos para la serie de escala diaria, transformaciones de datos, otro preprocesamiento que el alumno estime relevante para proporcionar la mejor predicción, etc.).
- Análisis inicial de cada una de las dos series: Qué se observa visualmente (tendencia o no, estacionalidad o no), justificando el análisis con datos objetivos (procedentes del análisis visual preliminar de la serie y sus componentes).
- En el caso de existir tendencia, justificar qué modelo de tendencia se utiliza para eliminarla (filtros, aproximación funcional, diferenciación). En el caso de no existir tendencia, justificar igualmente por qué motivo se ha llegado a esta conclusión.
- En el caso de existir estacionalidad, justificar qué modelo se utiliza para eliminarla. En el caso de no existir estacionalidad, justificar igualmente por qué motivo se ha llegado a esta conclusión.
- Explicación del procedimiento seguido para comprobar y conseguir la estacionariedad, en base a los ADF, ACF, PACF.



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada



- Justificar la selección del modelo de predicción.
- Explicar cómo se ha validado el modelo ajustado.
- Describir, en el caso de existir varios modelos de predicción, qué criterio se ha escogido para seleccionar el mejor de ellos (AIC, MSE, etc.), justificando la elección del criterio.
- Describir los pasos necesarios para conseguir la predicción real de los valores de la serie.

3. Condiciones de entrega

La entrega se realizará mediante la presentación en formato electrónico de una memoria de teoría y prácticas, junto con ficheros separados (scripts R) que resuelvan cada problema y los ficheros de datos de las series usadas listos para ser usados por los scripts de R. La práctica contribuirá a la calificación final de la asignatura en **4 puntos**, divididos entre **2 puntos** para la parte de teoría y **2 puntos** para la parte práctica.

La memoria tendrá 2 partes:

- **Parte teórica (2 puntos)**, donde se describirán los conceptos teóricos que se han utilizado para poder analizar las series de datos en cada punto de toma de decisiones, y el uso de cada uno para el análisis de series temporales (preprocesamiento, modelado de tendencia y estacionalidad, tests estadísticos, ACF y PACF, modelos AR/MA/ARIMA, etc.). Se organizarán en los siguientes apartados:
 - Preprocesamiento: Descripción de las técnicas utilizadas.
 - Análisis de tendencia y de estacionalidad: Descripción de las técnicas utilizadas para cada serie.
 - Estacionariedad: Técnicas para transformar una serie a estacionaria, y para analizar si es estacionaria o no.
 - Modelado de series temporales: Técnicas utilizadas
- **Parte práctica (2 puntos, 1 por cada serie)**, donde se describirá qué estación meteorológica se ha escogido y, paso a paso, el procedimiento seguido para analizar, modelar y predecir las dos series temporales. La memoria deberá tener, dentro de la parte práctica, subsecciones donde se analicen cada serie temporal por separado. Cada uno de los ítems comentados anteriormente en el apartado 2 de este guión tendrá la misma puntuación.

La memoria también deberá contener una portada con la siguiente información:

- Nombre del alumno



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada



- E-mail del alumno
- Nombre de la asignatura
- Nombre del Máster
- Texto “Trabajo autónomo I: Series Temporales”
- Fecha de entrega de la memoria.

Las soluciones a los problemas de predicción deberán entregarse a través de la plataforma PRADO, en la fecha indicada en la plataforma y anunciada con al menos una semana de antelación por el profesorado. En particular, se deberá subir un único fichero .zip que contenga:

- El PDF de la memoria.
- Dos ficheros de texto “**EstacionXXXX_diaria.txt**” y “**EstacionXXXX_mensual.txt**” con los datos de las series que se utilizarán por los scripts.
- Dos scripts R “**EstacionXXXX_diaria_DNI.R**” y “**EstacionXXXX_mensual_DNI.R**”, que contengan el **código comentado** con todos los pasos realizados para analizar la serie, donde **DNI** debe sustituirse por el DNI/NIE/pasaporte del estudiante.

Los ficheros script .R deberán estar **completamente comentados, describiendo cada paso que se realice**, justificando cada decisión tomada en el proceso de análisis, modelado y predicción. Además, forzosamente se deberán incluir comentarios al comienzo de cada script conteniendo la siguiente información:

- Línea 1: Nombre, apellidos y DNI del alumno
- Línea 2: E-mail del alumno
- Línea 3: Texto “Ejercicio de trabajo autónomo. Series temporales. Curso 20XX-20XX”, sustituyendo XX por el curso académico actual.

Tanto los ficheros como la versión electrónica en PDF de la memoria deberán remitirse por PRADO en la entrega habilitada para ello.