1 Origen de los datos

El conjunto de datos se ha creado a partir de observaciones meteorológicas procedentes de dos fuentes de información. Las observaciones comprenden los años 2014 al 2016 (3 años de mediciones).

Existe 1 medición cada 6 horas, por tanto son 4 mediciones al día, con lo que 365 días x 4 mediciones/dia = 1460 mediciones por año.

Por un lado, se han empleado mediciones obtenidas por una boya marítima, la cual está provista de diferentes sensores que le permiten realizar observaciones sobre diferentes variables meteorológicas (velocidad y dirección del viento, presión, temperatura del aire, etc.).

Por otro lado, se han utilizado datos matemáticos de reanálisis consistentes en variables meteorológicas (componentes U y V de la velocidad del viento, presión a nivel del mar, etc.), provenientes de un modelo de reanálisis cercano a la localización geográfica de la boya. Dichos datos se obtienen mediante un modelo de clima matemático.

Quizá algún intervalo no tenga todas esas mediciones porque no se pudo obtener por algún tipo de problema, por ejemplo rotura de la boya que hace las mediciones.

- Desde el patrón 1 al patrón 643, parte del año 2014.

- Desde el patrón 644 al patrón 2100, año 2015.

- Desde el patrón 2101 al patrón 3559, año 2016.

2 Descripción de los archivos disponibles

El conjunto de datos de está formado por 17 variables de entrada y 1 variable de salida.

Las variables de entrada son las siguientes:
Variable 1 (reanalisis) - air - Temperatura del aire en la superficie - Grados Kelvin
Variable 2 (reanalisis) - pres - Presión a nivel de superficie - Pascales
Variable 3 (reanalisis) - rhum - Humedad relativa a nivel de superficie - %
Variable 4 (reanalisis) - uwnd - Velocidad del viento de oeste a este (eje x) - Metros por segundo
Variable 5 (reanalisis) - vwnd - Velocidad del viento de sur a norte (eje y) - Metros por segundo
Variable 6 - WDIR - Dirección del viento en el sentido de las agujas del reloj - Grados
Variable 7 - WSPD - Velocidad media del viento - Metros por segundo
Variable 8 - GST - Velocidad de pico del viento - Metros por segundo
Variable 9 - DPD - Periodo de ola dominante - Segundos
Variable 10 - APD - Periodo medio de ola dominante - Segundos
Variable 11 - MWD - Dirección desde la que viene el periodo de ola dominante, DPD - Grados
Variable 12 - PRES - Presión a nivel de superficie - Hectopascales
Variable 13 - ATMP - Temperatura del aire en la parte superior de la boya - Grados Celsius
Variable 14 - WTMP - Temperatura a nivel del mar - Grados Celsius
Variable 15 - DEWP - Punto de rocio tomado a la misma altura que ATMP - Grados Celsius
Variable 16 - VIS - Visibilidad desde la boya - Millas náuticas
Variable 17 - TIDE - El nivel del agua en pies por encima o por debajo de la media inferior del agua - Pies

La variable de salida indica la altura de ola en las seis horas siguientes, la cual ha sido discretizada en 4 clases para una compresión y manejo de información más sencilla: {Baja, Media, Moderada, MuyAlta}.

Las variables de entrada están tomadas en el instante ***t*** y la salida hace referencia al instante ***t+1***, por lo tanto se trata de un **problema de predicción multiclase a 6 horas.**

3 Un conjunto de test para realizar alguna prueba sobre el preprocesamiento realizado

Dispone de un conjunto de *test* correspondiente a las mediciones realizadas en los años 2017 y 2018, al cual deberá realizar el mismo preprocesamiento, por si quiere comprobar el rendimiento obtenido al preprocesar el conjunto de datos.

Note que para *test* se pasan años completos, suele ser habitual esta práctica en el caso de series temporales. Imagine hacer un *hold-out* estratificado donde en entrenamiento caen solo meses de primavera y verano y en *test* los meses de invierno. El modelo debe ajustarse a datos de todos los meses, de ahí ese diseño experimental de años enteros.

Puede usar cualquier algoritmo de Weka, por ejemplo ***classifiers*** → ***functions*** → ***Logistic***

En la opción "*Supplied test set*" se indicaría el fichero del conjunto de test, mientras que el de entrenamiento corresponde al que se ha cargado desde la pestaña ***preprocess***