

Si quisieramos hacer una normalización mas justa a un conjunto de datos de test, se deberían usar los valores mínimos y máximo de train, que son los valores a partir de los cuales se ha entrenado el modelo.

A continuación se indica un ejemplo del proceso a seguir para normalizar el atributo "air". En primer lugar, se tienen que obtener los valores máximo y mínimo del atributo en train de manera precisa:

- Para ello, una vez abierto el conjunto de datos de train, hay que pulsar en la opción "Preprocess -> Edit..." y aparecerán todos los patrones con sus atributos.
- A continuación, se pulsa arriba (sobre el nombre del atributo "air") y de este modo se ordenarán sus valores. Para que se muestren todos los decimales hay que ampliar el tamaño de la celda (posicionar el cursor entre el atributo "air" y "pres" y cuando éste cambie pulsar y desplazar). De esta forma se obtiene que el valor mínimo es 272.499969 y el valor máximo es 289.200012 .

Para normalizar un conjunto de datos de test (una vez obtenidos el mínimo y máximo de train y se haya normalizado con el filtro correspondiente) se debe utilizar el filtro

`weka.filters.unsupervised.attribute.MathExpression:`

- En "expression" hay que insertar la siguiente ecuación para realizar la normalización:
$$(A-272.499969)/(289.200012-272.499969)$$
- En "ignoreRange" se indicará el índice del atributo: 1
- En "invertSelection" se seleccionará True para que sólo realice la normalización del atributo indicado.
- Finalmente, se pulsará en "OK" para confirmar la configuración y en "Apply" para aplicar el filtro.

Este proceso tendría que realizarse con los restantes atributos, indicando para cada uno sus correspondientes valores máximo y mínimo y el índice que tiene en el conjunto de datos.