

Entrega: subir a moodle, antes del domingo 28 de abril a las 23:59 horas, un documento PDF en el que se incluya, para cada pregunta:

1. El enunciado de la pregunta.
2. Los comandos de R utilizados para responder a la pregunta.
3. Los resultados mostrados por R como respuesta a cada comando.
4. Copia, en su caso, de las gráficas generadas por R.
5. Cuando se pida, el análisis de los resultados obtenidos.

Valoración: 100% de la nota final.

---

### 1. Clustering

Tenemos los datos de vinos de España de distintas denominaciones de origen, se requiere hacer una segmentación de vinos españoles en función de precio, calidad y acidez del vino.

- (a) Instalar el paquete *Tidyverse*.
- (b) Cargar el fichero en un *dataframe*, considerando el tipo de separador adecuado.
- (c) Identificar la estructura de los datos y realizar la coerción de los datos que no se encuentren en formato numérico y que sean necesarios para la segmentación.
- (d) Eliminar aquellas filas que no contengan datos o que no estén disponibles.
- (e) Valorar si es necesario realizar un escalado de algunos de los datos: *price*, *rating*, *acidity*.
- (f) Haz un *clustering* de todos los vinos. Considera si puedes ponderar o escalar los datos de alguna manera razonable. Usa el método *elbow* para determinar el valor más apropiado de *k*.
- (g) Repite el proceso solo para los vinos de denominación "Rioja".
- (h) Estudia las características de los clústeres obtenidos y comenta las observaciones y conclusiones que se pueden extraer.

## 1. Series temporales

Tenemos los datos del precio de venta del excedente de sistemas fotovoltaicos a la red eléctrica nacional, para distintos periodos de tiempo y con distinta resolución temporal.

### Datos semanales

- (a) Cargue el archivo de datos 'precios\_excedente\_peninsula\_semana.xls' y guárdelo en un dataframe.
- (b) ¿Cuántos días de datos hay disponibles? ¿Con qué resolución? Guarde un 20% de los mismos para test, y el resto como datos de entrenamiento.
- (c) Ajuste un modelo ARIMA posiblemente estacional (¿periodo?) de forma manual, justificando los parámetros escogidos. Realice también un ajuste con autoarima sin aproximación.
- (d) Analice el ACF, el PACF y la normalidad de los residuos para ambos modelos. Comente si hay diferencias significativas.
- (e) Utilice ambos modelos para predecir un día de datos y compárelos con los datos guardados para test. ¿Qué modelo es más preciso? ¿Cuál tiene más incertidumbre en las predicciones?

### Datos mensuales

- (a) Cargue el archivo de datos 'precios\_excedente\_peninsula\_mes.xls' y guárdelo en un dataframe.
- (b) ¿Cuántos días de datos hay disponibles? ¿Con qué resolución? Guarde cuatro semanas de datos para test, y el resto como datos de entrenamiento.
- (c) Tras una primera diferenciación sin estacionalidad, compare la serie resultante de aplicar una segunda diferenciación con estacionalidad si se considera un patrón diario o un patrón semanal. ¿Cuál es más adecuado? Comente qué ocurre con los precios el fin de semana, y por qué esto hace más conveniente un patrón u otro.
- (d) Realice un ajuste con autoarima con aproximación, y utilice el modelo resultante para predecir 3 días de datos. Comente los resultados.