

**Fundamentos de Aprendizaje Automático - 2025 – Grado en Inteligencia Artificial****Ejercicios prácticos**

Escribe un script de R con los comandos necesarios para resolver los siguientes ejercicios que será enviado por el campus virtual donde también está el material necesario. Usa # al principio de cada línea para incluir tus comentarios sobre los resultados obtenidos. La valoración estará centrada en la precisión de los comentarios sobre las elecciones tomadas, los resultados obtenidos y en su interpretación estadística.

Para personalizar alguna de las preguntas del examen, crea el objeto `midni=xxxx` (en xxxx van las últimas 4 cifras de tu DNI). Ejecuta los siguientes comandos como los iniciales en tu script: `set.seed(midni), A1= midni %% 2, A2= sample(1:25,3), load("KorTemp.RData")` y `lmues=sample(1:nrow(KorTemp),300)`

El data.frame `KorTemp` que se carga en memoria con el comando `load("KorTemp.RData")` (con el fichero descargado en el directorio de trabajo) contiene variables de la predicción meteorológica en los meses de verano en estaciones de Korea del Sur. Las variables que empiezan por L son medidas del último período disponible y las que empiezan por F son medidas del período futuro. Las variables son: `Est`: Estación (1–25)(factor), `LTMax`, `LTmin`: Temp. Máx y Mín. `LRHmax`, `LRHmin`: Humedad relativa Máx. y Mín., `LTMax_D`, `LTmin_D`: Temp. Máx y Mín anterior, `LWS`: Velocidad de viento, `LLH`: Flujo de calor, `LRadSol`: Radiación Solar, `FTMax`, `FTmin`: Temp. Máx. y Mín. futuro, `mes`: mes (factor) y `year`: Año (factor).

1. (2.5 pts) Se desea predecir la **temperatura máxima futura** (`FTMax`), en función de variables meteorológicas del último periodo. Ajusta por el método de mínimos cuadrados el siguiente modelo de regresión múltiple:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$  donde:  $X_1 = \text{LWS}$ ;  $X_2 = \text{LRadSol}$ ;  $X_3 = \text{LLH}$ ;  $X_4 = \text{LTMax}$ ;  $X_5 = \text{LRHmax}$  y  $X_6 = \text{LTmin}$ 
  - a) Obtén los coeficientes estimados del modelo, así como intervalos de confianza al 95 % para los coeficientes del modelo. Calcula los valores ajustados del modelo y calcula la suma residual de cuadrados.
  - b) Calcula el valor del coeficiente de correlación de Pearson entre la variable respuesta  $Y$  y el predictor  $X_1$ .
  - c) Define y calcula el coeficiente de correlación **parcial** entre  $X_1$  y  $X_2$ , controlando por el resto de las variables explicativas.
  - d) Considera un **modelo reducido** que contiene solo un subconjunto de las variables explicativas incluidas en el modelo completo. Formula y contrasta, mediante un **test t**, la hipótesis nula de que los coeficientes asociados a las variables excluidas son iguales a cero.

- e) Compara el modelo completo con el modelo reducido mediante un **test**  $F$ . Expón claramente la hipótesis nula, calcula el estadístico de contraste, determina el valor crítico correspondiente y extrae las conclusiones en función del resultado del test.
- f) Compara el coeficiente de determinación ajustado para ambos modelos.

Compara todos los resultados que obtienes con sus versiones manuales, verifica que coincidan e interpreta los resultados.

2. (1.5 pts) Calcula las componentes principales de las variables del último período disponible. Dibuja las puntuaciones de las dos primeras componentes respecto al mes (`mes`). Interpreta las componentes que expliquen conjuntamente más del 75 % de la variabilidad.
3. (1 pts) Calcula un modelo de regresión para explicar ( $A1=0:FTMax$ ,  $A1=1:FTmin$ ) con las variables del último período disponible (empiezan por `L`), con el objetivo de usar el mejor modelo con el menor número de covariables. Explica el proceso seguido, las elecciones tomadas y los resultados obtenidos.
4. (1 pts) Seleccionando las filas del conjunto de datos dada por `lmues: KorTemp[lmues,]`, estimar un modelo de regresión no lineal para explicar ( $A1=0:FTMax$ ,  $A1=1:FTmin$ ) usando la variable `LRadSol`. Comenta las opciones elegidas y dibuja el modelo de regresión obtenido sobre los datos originales.
5. (1 pts). Para las estaciones seleccionadas en `A2`, elaborar dos modelos de clasificación (uno basado en Regla de Bayes y el otro en técnicas de regresión) con las variables `LTHmax`, `LRHmax`, `LWS` y `LLH` que intente clasificar los datos por estación. Justificar las elecciones tomadas y estimar el error de mala clasificación que tendríamos si aplicásemos los modelos a nuevos datos.