



MAESTRÍA EN ESTADÍSTICA Y CIENCIA DE DATOS

## **Práctica de optimización**

SEMINARIO II  
CURSO 2025-2026

## Instrucciones de la práctica

La práctica se realiza en grupos de 2 ó 3 personas.

- **Fecha límite entrega:** 30 de noviembre de 2025 a las 23:59.
- **Lugar:** Campus Virtual, sección de Evaluaciones. Se entregará un archivo zip con el siguiente nombre "Optimizacion\_grupoX.zip" siendo X la letra de cada grupo.
- **Formato:** Archivo pdf y script/notebook de Python. Será obligatoria la entrega de un documento pdf en el que se muestren todos los resultados y su análisis. Este documento deberá contener el nombre de los participantes, el grupo y los correspondientes enunciados para que el documento tenga contexto.

## Enunciado de la práctica

El Observatorio de Turismo Sostenible de Mallorca está interesado en desarrollar un índice que refleje la sostenibilidad del turismo en la zona. Para ello, ha recopilado datos históricos del IBESTAT (Instituto de Estadística de las Islas Baleares) que distintos expertos en turismo y sostenibilidad han identificado como fundamentales. El conjunto de datos, disponible en el Aula Virtual bajo el nombre Mallorca.csv, contiene las siguientes variables:

- $x_1$ : SchoolingRate - Tasa de escolarización a 5 años en ambos géneros
- $x_2$ : Poverty - Población en riesgo de pobreza relativa (ingresos < 60 % mediana)
- $x_3$ : UnaccountedWater - Agua no contabilizada
- $x_4$ : OpenEstablishments - Establecimientos abiertos estimados por período, isla y categoría
- $x_5$ : RenewableResources - Cantidad anual de energía consumida a partir de recursos renovables (Mwh) en comparación con el consumo mundial de energía a nivel objetivo por año
- $x_6$ : PassengersArriving - Pasajeros que llegan a la isla por mar y aire
- $x_7$ : MaritimeTraffic - Tráfico de pasajeros en líneas marítimas regulare.
- $x_8$ : VehicleRegistration - Número de matriculaciones de vehículos en Baleares
- $y$  - Índice de turismo sostenible calculado por expertos
- year - Año de referencia

El objetivo de la práctica es determinar qué variables tienen mayor impacto en la estimación del índice de turismo sostenible. Para ello, se deben ajustar distintas rectas de regresión lineal con el propósito de encontrar el mejor modelo para predecir y se deben interpretar los resultados.

La práctica se realizará en Python, utilizando la librería PuLP o Gurobipy para la parte de optimización. En todo momento se valorará que el código sea claro y conciso. Todas las respuestas deben estar debidamente justificadas.

## Primera parte

Dado el conjunto de datos previo:

1. (1 pto) Realice un análisis exploratorio de los mismos

Una vez realizado dicho análisis exploratorio y comprendidos los datos y las variables, obtenga la recta que ajusta la variable y en función del resto de variables explicativas bajo los siguientes criterios:

2. (1 pto) Mediante programación lineal minimizando la suma de las desviaciones absolutas (norma  $\ell_1$ ). Debe plantearse el problema de programación lineal de forma teórica y mostrar su resolución con Python. Se debe escribir la ecuación de la recta resultante y el error de la misma.
3. (1 pto) Mediante programación lineal minimizando la desviación absoluta máxima (norma  $\ell_\infty$ ). Debe plantearse el problema de programación lineal de forma teórica y mostrar su resolución con Python. Se debe escribir la ecuación de la recta resultante y el error de la misma.
4. (1 pto) Mediante el método de mínimos cuadrados usando la función `LinearRegression()` de la librería `scikit-learn` de Python (norma  $\ell_2$ ). Se debe escribir la ecuación de la recta resultante y el error de la misma.
5. (1 pto) Enfrente gráficamente el target y las predicciones obtenidas por los tres métodos anteriores. Analice los resultados.

## Segunda parte

Desde el Observatorio de Turismo de Mallorca quieren identificar qué cambios podrían implementarse cada año para mejorar el índice de turismo sostenible. Para ello, el equipo de científicos de datos propone utilizar técnicas de contrafacticos basadas en optimización lineal. En particular, se busca determinar los mínimos cambios posibles en las variables explicativas que permitan alcanzar un valor deseado en la variable respuesta y (se denotará por  $\varepsilon$  el aumento deseado). Se asume que el índice de turismo sostenible sigue el modelo:

$$y = \sum_{j=1}^p w_j x_j$$

donde  $x_j$  representa cada variable explicativa y  $w_j$  su peso correspondiente. Se imponen las siguientes condiciones:

- Los pesos están normalizados:  $\sum_{j=1}^p w_j = 1$
- Las variables explicativas están acotadas en el intervalo  $[0, 1]$ .

A partir de esto, se plantean las siguientes cuestiones:

6. (2 pts) Plantee el problema de programación lineal asociada a esta optimización contrafáctica. Recuerde que el objetivo es minimizar tanto la magnitud de los cambios en las variables explicativas como el número total de modificaciones necesarias para lograr el valor deseado de  $y$ . El problema además tiene algunas restricciones asociadas:
  - Selección de variables: Se permite modificar como mínimo  $\mu_l$  variables explicativas y como máximo  $\mu_u$ .
  - Cotas en los cambios: El cambio en cada variable explicativa  $x_j$ , denotado como  $\beta_j$ , está acotado superiormente por una función de su desviación estándar  $f_{std}$ , es decir,  $\beta_j \leq a f_{std}(x_j)$ , siendo  $a$  una constante.
  - Rango de las variables: Las variables explicativas deben permanecer en su rango  $[0, 1]$ .

Para los siguientes apartados, asúmase que  $w_j = 1/p, \forall j = 1, \dots, p$ . Se aclara que el valor de la constante  $a$  queda a la elección y justificación del alumno.

7. (1.5 pto) ¿Qué cambios se deben realizar en las variables explicativas para lograr un incremento del 1 % en el valor del índice de turismo en el año 2024? ¿Y un incremento de un 5 % si como máximo se pueden modificar 4 variables?
8. (1 pto) Si el gobierno de las Islas Baleares solo puede invertir dinero en una dimensión de las estudiadas, ¿cuál le recomendaría modificar para lograr un incremento del 25 % en 2024? Justifique su respuesta.
10. (0.5 pts) Formule las restricciones adicionales necesarias para modelar las siguientes condiciones:
  - a) Si se modifica la variable `MaritimeTraffic`, también debe modificarse la variable `PassengersArriving`.
  - b) Se debe modificar una y solo una entre `SchoolRate` y `RenewableResources`.
  - c) Al menos una de las variables `Poverty` o `VehicleRegistration` debe ser modificada.