

## Tarea 1: Aplicaciones de la Matemática en Ingeniería

28 de agosto de 2023

Considere un conjunto de datos que contiene valores para seis características biomecánicas utilizadas para clasificar a los pacientes ortopédicos en 2 clases (normal o anormal) (descargar desde AULA el archivo `column_2C.txt`). Más precisamente, se tienen las siguientes covariables:

- †  $X_1$ : incidencia pélvica
- †  $X_2$ : inclinación pélvica
- †  $X_3$ : ángulo de lordosis lumbar
- †  $X_4$ : pendiente sacra
- †  $X_5$ : radio pélvico
- †  $X_6$ : grado de espondilolistesis.

La variable respuesta ( $Y$ ) es una etiqueta, "NO" o "AB", que indica si el paciente es normal o anormal, respectivamente.

Responda las siguientes preguntas:

- a. ¿Cuántas observaciones hay en el conjunto de datos? ¿Cuántas observaciones corresponden a cada etiqueta?
- b. Realice histogramas para cada una de las covariables (separando las observaciones de acuerdo a las distintas etiquetas). Además, reporte medidas de tendencia central, dispersión y forma. Comente los resultados.
- c. Para diferentes parejas de covariables, reporte los gráficos de dispersión (es decir, grafique  $X_i$  versus  $X_j$ ) usando diferentes colores/símbolos de acuerdo a las distintas etiquetas. Comente los patrones que observa en estos gráficos.
- d. Ajuste un modelo de regresión logística con todas las covariables. Reporte y comente los resultados obtenidos.
- e. Con el modelo ajustado en el inciso anterior prediga la etiqueta de un paciente con las siguientes características:  $(X_1, X_2, X_3, X_4, X_5, X_6) = (60, 20, 50, 50, 100, 10)$ .
- f. ¿Es razonable disminuir el número de covariables? Justifique su análisis rigurosamente.
- g. ¿Es adecuado utilizar el método de Bayes ingenuo en este conjunto de datos? Justifique su respuesta.
- h. Independientemente de la respuesta del inciso anterior, ajuste un modelo de Bayes ingenuo. Reporte las funciones de densidad estimadas. Comente sus resultados.
- i. Ajuste un árbol de clasificación. Reporte y comente sus resultados.
- j. Use los ajustes de los incisos **h.** y **i.** para predecir la etiqueta de un paciente con las mismas características que en el inciso **e.** ¿La conclusión coincide con la regresión logística?