

# Examen Parcial 1 - Estadística Aplicada III

Diana Isabel Muñoz Castillo

October 4, 2024

## Pregunta 1 - Wine Data

Dividí los datos en un 80% para entrenamiento y 20% para prueba. Entrené tres clasificadores: Análisis Discriminante Lineal (LDA), k-Nearest Neighbors (k-NN) y **Regresión Logística Multinomial**. Finalmente, seleccioné la **Regresión Logística Multinomial** debido a su flexibilidad y a que no asume normalidad en los datos.

### 1. Elección del Clasificador

La **Regresión Logística Multinomial** fue elegida porque, al ser un modelo paramétrico, permite interpretar la probabilidad de que una instancia pertenezca a cada clase. A diferencia de LDA (precisión: 93.10%), que requiere que las clases sigan distribuciones normales, la **Regresión Logística Multinomial** (precisión: 100%) es más flexible y adecuada para datos donde estas suposiciones pueden no cumplirse. Aunque k-NN (precisión: 93.10%) es una buena opción, su capacidad de interpretación es limitada y la elección del valor de  $k$  influye mucho en el rendimiento, lo que lo hace menos robusto en este caso.

### 2. Evaluación del Clasificador

Evalué el rendimiento de la **Regresión Logística Multinomial** con dos criterios: **precisión** y **matriz de confusión**. La precisión fue del 100%, indicando que el modelo clasificó correctamente todas las instancias del conjunto de prueba. La matriz de confusión mostró que no hubo errores de clasificación entre las clases, lo que refleja un alto rendimiento (ver Figura 1).

Este resultado perfecto sugiere que el modelo se ajusta bien a este conjunto de datos. Sin embargo, recomiendo incrementar el número de observaciones para asegurar que no exista sobreajuste.

### 3. Explicación del Clasificador

La **Regresión Logística Multinomial** asigna probabilidades a cada clase, lo que facilita interpretar la confianza en las predicciones. El análisis de la distribución de probabilidades mostró que el modelo asigna valores cercanos a 1 para la clase correcta en la mayoría de los casos, lo que indica alta seguridad en sus predicciones (ver Figura 2).

## Anexos

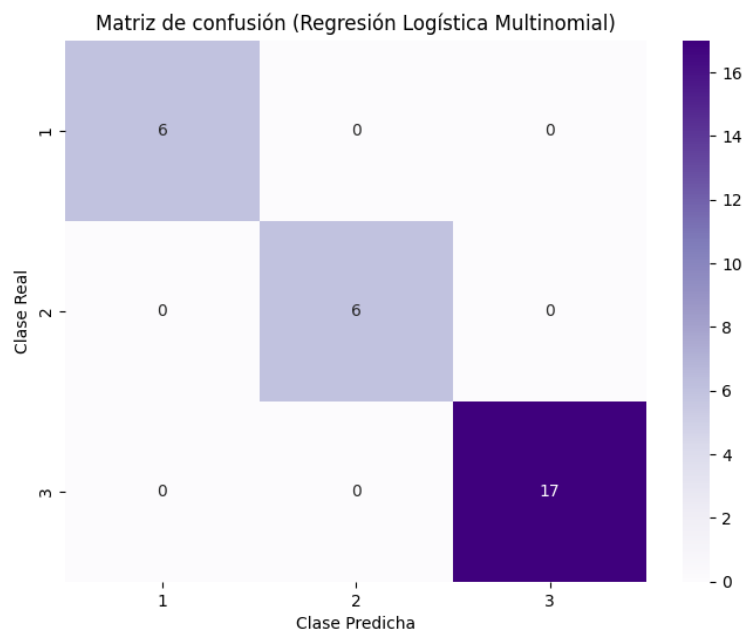


Figure 1: Matriz de Confusión - Regresión Logística Multinomial

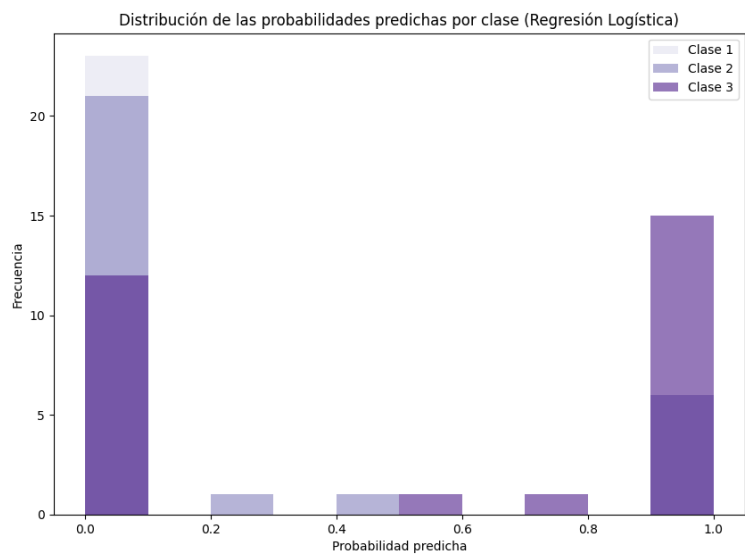


Figure 2: Distribución de las Probabilidades Predichas por Clase

## Referencias

- GitHub CRAN MMST - Wine Dataset. Recuperado de: <https://github.com/cran/MMST/blob/master/data/wine.rda>.