

Task 9

Tema: Clasificación - Modelo Logit (Predicción de Compra)

Profesor: Eduard F. Martínez González

Octubre 2025

Introducción

El presente taller tiene como propósito aplicar técnicas de **clasificación supervisada** mediante el **Modelo Logit**, utilizando un conjunto de datos simulado que representa la probabilidad de que un cliente *compre* o no un producto determinado.

Cada observación corresponde a un cliente, con información sobre su edad, ingresos, comportamiento digital y relación con la marca.

El objetivo del análisis es estimar la **probabilidad de compra** y evaluar qué factores influyen en dicha decisión, aplicando el modelo logístico e interpretando sus resultados.

Este tipo de modelo permite a las empresas comprender los determinantes del comportamiento del consumidor y diseñar estrategias de marketing basadas en datos.

Instrucciones

- El conjunto de datos del taller se encuentra disponible en el siguiente enlace.
- La entrega debe incluir:
 - Un archivo en formato .R con el desarrollo completo del taller.
 - Un archivo en formato PDF con las interpretaciones, tablas y gráficas solicitadas.
- El script en R debe contener:
 - Nombres de los integrantes y versión de R utilizada.
 - Carga de librerías necesarias (`tidyverse`, `broom`, `PROC`, etc.).
 - Código organizado y documentado con comentarios: `## Punto 1`, `## Punto 2`, etc.

```
## Nombre(s) de Autor(es)
## R version 4.5.0
```

```

rm(list = ls()) # limpiar entorno

## cargar librerias
library(tidyverse)
library(broom)
library(pROC)

# cargar datos
df <- read.csv("datos_compra_logit.csv")

# vista previa
head(df)

# estructura general
str(df)

# modelo logit
modelo <- glm(compra ~ edad + ingreso_mensual +
  visitas_web +
  tiempo_en_sitio + anuncios_vistos +
  descuento_recibido + miembro_programa,
  data = df, family = binomial(link = "logit"))

summary(modelo)

```

Taller: Clasificación – Modelo Logit (Predicción de Compra)

1. (15 pts) Revise la base de datos y describa:
 - El número total de observaciones y variables.
 - La proporción de clientes que compraron frente a los que no compraron.
 - El comportamiento general de las variables mediante gráficos exploratorios.
2. (20 pts) Estime un **Modelo Logit** con la variable dependiente **compra**. Interprete los coeficientes obtenidos e identifique qué variables aumentan o reducen la probabilidad de compra.
3. (15 pts) Calcule las probabilidades predichas de compra para cada cliente y clasifique a los clientes como compradores (1) o no compradores (0) usando un umbral de 0.5.
4. (15 pts) Construya la **matriz de confusión** y calcule las métricas de desempeño:
 - Exactitud (Accuracy)
 - Sensibilidad (Recall)

- Especificidad
 - Precisión (Precision)
5. (15 pts) Calcule el **Área Bajo la Curva (AUC)** y grafique la **Curva ROC** usando el paquete **pROC**. Interprete el valor del AUC en términos de capacidad predictiva.
 6. (10 pts) Visualice la relación entre la probabilidad predicha de compra y el tiempo promedio en el sitio web. Describa la forma de la relación y su interpretación práctica.
 7. (10 pts) Interprete los hallazgos más relevantes:
 - ¿Qué variables son los mejores predictores de compra?
 - ¿Qué tan bien clasifica el modelo?
 - ¿Qué implicaciones tienen los resultados para estrategias de marketing o segmentación?

Total: 100 pts