



Universidad Nacional de Ingeniería  
Facultad de Ciencias  
Escuela Profesional de Ciencia de la Computación

## Práctica Calificada II Minería de Datos

CC442

23/01/2026

Ciclo: 2025-III

**Puntaje:** 20 puntos

**Duración.** 120 minutos

### PARTE I (10 puntos)

#### Condición Financiera de los Bancos

El archivo **Banks.csv** incluye datos sobre una muestra de 20 bancos. La columna **Financial Condition** (Condición Financiera) registra el juicio de un experto sobre la condición financiera de cada banco. Esta variable de resultado toma uno de dos valores posibles: **débil (weak)** o **fuerte (strong)**, según la condición financiera del banco.

Los predictores son dos ratios utilizados en el análisis financiero de los bancos:

- **TotLnsLses\_Assets:** es la proporción de préstamos y arrendamientos totales frente a los activos totales.
- **TotExp\_Assets:** es la proporción de gastos totales frente a los activos totales.

El objetivo es utilizar los dos ratios para clasificar la condición financiera de un nuevo banco.

Ejecute un modelo de regresión logística (sobre todo el conjunto de datos) que modele el estado de un banco en función de las dos medidas financieras proporcionadas. Especifique la clase de éxito como **débil (weak)** (esto es similar a crear una variable binaria/dummy que sea 1 para los bancos financieramente débiles y 0 en caso contrario), y utilice el valor de corte (cutoff) predeterminado de 0.5.

a. Escriba la ecuación estimada que asocia la condición financiera de un banco con sus dos predictores en tres formatos:

1. El **logit** como función de los predictores.
2. Los **odds** (posibilidades) como función de los predictores.
3. La **probabilidad** como función de los predictores.

b. Considere un nuevo banco cuyo ratio **TotLnsLses\_Assets = 0.6** y ratio **TotExp\_Assets = 0.11**. A partir de su modelo de regresión logística, estime las siguientes cuatro cantidades para este banco (utilice Python para realizar todos los cálculos intermedios; muestre sus respuestas finales con cuatro decimales):

- El logit.

- Los odds.
- La probabilidad de ser financieramente débil.
- La clasificación del banco (utilice un punto de corte = 0.5).

c. El valor de corte de 0.5 se utiliza en conjunto con la probabilidad de ser financieramente débil. Calcule el **umbral (threshold)** que debería usarse si quisiéramos realizar una clasificación basada en los **odds** de ser financieramente débil, y el umbral para el logit correspondiente.

d. Interprete el coeficiente estimado para el ratio de préstamos y arrendamientos totales sobre activos totales (TotLns&Lses/Assets) en términos de los odds de ser financieramente débil.

e. Cuando un banco que está en una condición financiera deficiente se clasifica erróneamente como financieramente fuerte, el costo de esta clasificación errónea es mucho mayor que cuando un banco financieramente fuerte se clasifica erróneamente como débil. Para minimizar el costo esperado de una clasificación errónea, ¿debería aumentarse o disminuirse el valor de corte (cutoff) para la clasificación (que actualmente es de 0.5)?

## PARTE II (10 puntos)

### Instrucciones para el alumnado.

Realiza los cálculos claros y ordenados.

No uses librerías de ML; puedes usar calculadora simple.

### — Descenso por gradiente (logística + L2, 1 iteración)

**Modelo.**  $p_\theta(y = 1 | \mathbf{x}) = \sigma(z)$ , con  $z = w_0 + w_1x_1 + w_2x_2$  y  $\sigma(t) = \frac{1}{1+e^{-t}}$ .  
**Pérdida media regularizada.**

$$\mathcal{L}(\mathbf{w}) = \frac{1}{m} \sum_{i=1}^m \left[ -y_i \ln p_i - (1 - y_i) \ln(1 - p_i) \right] + \frac{\lambda}{2m} (w_1^2 + w_2^2)$$

(No penalices el sesgo  $w_0$ ).

**Datos:**  $(x_1, x_2, y) = \{(0, 1, 0), (1, 0, 0), (1, 2, 1), (2, 1, 1), (0, 2, 1), (2, 2, 0)\}$ ,  $m = 6$ .  
**Inicio:**  $w_0^{(0)} = 0$ ,  $w_1^{(0)} = 1$ ,  $w_2^{(0)} = 0$ .  $\eta = 1$ ,  $\lambda = 1$ .

Gradientes a calcular (promedio).

$$\left[ \frac{\partial \mathcal{L}}{\partial w_0}; \frac{\partial \mathcal{L}}{\partial w_1}; \frac{\partial \mathcal{L}}{\partial w_2} \right]$$

Tablas de trabajo (Iteración 1 con  $w^{(0)}$ ).

$i$	$x_{1,i}$	$x_{2,i}$	$y_i$	$z_i = w_0 + w_1 x_{1,i} + w_2 x_{2,i}$	$p_i = \sigma(z_i)$
1	0	1	0		
2	1	0	0		
3	1	2	1		
4	2	1	1		
5	0	2	1		
6	2	2	0		

Gradientes y actualización.

$$\left. \begin{aligned} \frac{\partial \mathcal{L}}{\partial w_0} \Big|_{(0)} &= \\ \frac{\partial \mathcal{L}}{\partial w_1} \Big|_{(0)} &= \\ \frac{\partial \mathcal{L}}{\partial w_2} \Big|_{(0)} &= \\ w_0^{(1)} &= w_0^{(0)} - \eta \frac{\partial \mathcal{L}}{\partial w_0} \Big|_{(0)}, \quad w_1^{(1)} = w_1^{(0)} - \eta \frac{\partial \mathcal{L}}{\partial w_1} \Big|_{(0)}, \quad w_2^{(1)} = w_2^{(0)} - \eta \frac{\partial \mathcal{L}}{\partial w_2} \Big|_{(0)} \end{aligned} \right\}$$

Error final. Calcula  $\mathcal{L}(\mathbf{w}^{(1)})$  usando los nuevos pesos actualizados.

Pista. Puedes redondear  $\sigma(t)$  y  $\ln(\cdot)$  a 3 decimales para facilitar el cálculo a mano.