Ejercicio Regresiones

Inteligencia Artificial

Pronóstico de Compra de Hardware

Una empresa de tecnología necesita predecir si comprará nuevo equipo de cómputo el próximo mes con base en su histórico.

Si la predicción es incorrecta y no se adquiere el hardware a tiempo, la empresa podría enfrentar interrupciones operativas debido a que sus colaboradores no podrán trabajar.

Evidencia

Se te proporciona el siguiente conjunto de datos y se te pide aplicar **dos iteraciones de gradiente descendente** para actualizar los parámetros del modelo.

Años desde la última actualización	Compra de PC
1	No
2	No
3	Sí
4	Sí

¿Qué tipo de Regresión entrenar?

Ejercicio de Regresión Logística

Para ayudar en esta decisión, se necesita elaborar un modelo de **regresión logística** que estima la probabilidad de que la empresa realice una compra.

Es fundamental que el modelo sea preciso, ya que una mala predicción podría resultar en compras innecesarias (gasto excesivo) o en la falta de equipos cuando realmente se necesitan, afectando la operación del negocio (Reducir los falsos negativos – Más vale prevenir que lamentar!).

Ejercicio de Regresión Logística

Se utilizará **regresión logística** para modelar la probabilidad de compra en función del tiempo desde la última actualización. Para ello, se implementará **dos iteraciones** del algoritmo de descenso de gradiente para actualizar los parámetros B0 y B1 y una tasa de aprendizaje de 0.10

- Función sigmoide (probabilidad de que sí se compre la PC):

$$p_i = \frac{1}{1 + e^{-z}}$$

- Ecuación del modelo:

$$P(y = 1|x) = p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

Pasos a Seguir

1. Calcular la probabilidad de que y sea 1 dadas la(s) características, es decir, P(y=1|x) para cada dato.
2. Calcular la función de costo $J(\beta)$.
3. Actualizar los valores de β_0 y β_1 usando la regla del descenso de gradiente.
4. Repetir el proceso.

X	Y
1	0
2	0
3	1
4	1

Conjunto de Datos

Iteración 0

- Parámetros iniciales: $\beta_0^{(0)} = -2.400, \quad \beta_1^{(0)} = 0.897$
- Tasa de aprendizaje: $\alpha = 0.1$

Función para calcular probabilidades:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

Calcular el costo:

$$J = -\frac{1}{4}\sum_{i=1}^{4}[y_i ln(P_i) + (1 - y_i) ln(1 - P_i)]$$

Primera Iteración

Calculo de probabilidades para cada x:

- Para x=1:

$$P_1 = \frac{1}{1+e^{-(-2.400+0.897\cdot 1)}} = \frac{1}{1+e^{(1.503)}} \approx 0.1819\ldots$$

- Para x=2:

$$P_2 = \frac{1}{1+e^{-(-2.400+0.897\cdot 2)}} = \frac{1}{1+e^{(0.606)}} \approx 0.3529\ldots$$

- Para x=3:

$$P_3 = \frac{1}{1+e^{-(-2.400+0.897\cdot 3)}} = \frac{1}{1+e^{-(0.291)}} \approx 0.5722\ldots$$

- Para x=4:

$$P_4 = \frac{1}{1+e^{-(-2.400+0.897\cdot 4)}} = \frac{1}{1+e^{-(1.188)}} \approx 0.7663\ldots$$

Primera Iteración

Calcular el costo:

$$J = -\frac{1}{4}\sum_{i=1}^{4}[y_i ln(P_i) + (1 - y_i) ln(1 - P_i)]$$

Primera Iteración

Calcular el costo:

$$J = -\frac{1}{4}[0 \cdot ln(0.1819) + 0 \cdot ln(0.3529) + 1 \cdot ln(0.5722) + 1 \cdot ln(0.7663) + (1-0) \cdot ln(1-0.1819) + (1-0) \cdot ln(1-0.3529) + (1-1) \cdot ln(1-0.5722) + (1-1) \cdot ln(1-0.7663)]$$

$$= -\frac{1}{4}[0 + 0 + (-0.5583) + (-0.2662) + (-0.2008) + (-0.4353) + 0 + 0]$$

$$= -\frac{1}{4}[-1.4606]$$

$$= \frac{1.4606}{4}$$

$$\approx 0.3651$$

Primera Iteración

Fórmula de actualización de los parámetros:

- Para β_0:

$$\beta_0 \coloneqq \beta_0 - \alpha \cdot \frac{1}{m}\sum_{i=1}^{m}\left(P_i - y^{(i)}\right)$$

- Para β_1:

$$\beta_1 \coloneqq \beta_1 - \alpha \cdot \frac{1}{m}\sum_{i=1}^{m}\left(P_i - y^{(i)}\right)x^{(i)}$$

Primera Iteración

β_0:

$$Gradiente = \frac{1}{4}[0.1819 + 0.3529 - 0.4278 - 0.2337] \approx -0.03167 \ldots$$

β_1:

$$Gradiente = \frac{1}{4}[0.1819 \cdot 1 + 0.3529 \cdot 2 - 0.4278 \cdot 3 - 0.2337 \cdot 4] \approx -0.33245 \ldots$$

Primera Iteración

Actualización de parámetros:

- $\beta_0^{(1)} = -2.400 - 0.1 \cdot (-0.0317) \approx -2.400 + 0.00317 = -2.3968 \ldots$
- $\beta_1^{(1)} = 0.897 - 0.1 \cdot (-0.3325) \approx 0.897 + 0.03325 = 0.9302 \ldots$

Segunda Iteración

Nuevos Parámetros:

$$\beta_0 = -2.3968$$
$$\beta_1 = 0.9302$$

Se sustituyen en la ecuación sigmoide:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

Segunda Iteración

Calculo de probabilidades para cada *x*:

- Para x=1:

$$\mathrm{P}_1 = \frac{1}{1 + e^{-(-2.3968 + 0.9302 \cdot 1)}} = \frac{1}{1 + e^{-(-1.4666)}} \approx 0.1874 \ldots$$

- Para x=2:

$$P_2 = \frac{1}{1 + e^{-(-2.3968 + 0.9302 \cdot 2)}} = \frac{1}{1 + e^{-(-0.5364)}} \approx 0.3690 \ldots$$

- Para x=3:

$$P_3 = \frac{1}{1 + e^{-(-2.3968 + 0.9302 \cdot 3)}} = \frac{1}{1 + e^{-(0.3938)}} \approx 0.5972 \ldots$$

- Para x=4:

$$P_4 = \frac{1}{1 + e^{-(-2.3968 + 0.9302 \cdot 4)}} = \frac{1}{1 + e^{-(1.3248)}} \approx 0.7898 \ldots$$

Segunda Iteración

Calcular el costo:

$$J = -\frac{1}{4}[0 \cdot ln(0.1874) + 0 \cdot ln(0.3690) + 1 \cdot ln(0.5972) + 1 \cdot ln(0.7898) + (1-0) \cdot ln(1-0.1874) + (1-0) \cdot ln(1-0.3690) + (1-1) \cdot ln(1-0.5972) + (1-1) \cdot ln(1-0.7898)]$$

$$= -\frac{1}{4}[-1.4194]$$

$$= \frac{1.4194}{4} \approx 0.3548 \ldots$$

Segunda Iteración

Fórmula de actualización de los parámetros:

- Para β_0:

$$\beta_0 \coloneqq \beta_0 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \left(P_i - y^{(i)}\right)$$

- Para β_1:

$$\beta_1 \coloneqq \beta_1 - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} \left(P_i - y^{(i)}\right) x^{(i)}$$

Segunda Iteración

β_0:

$$Gradiente = \frac{1}{4}[0.1874 + 0.3690 - 0.4028 - 0.2102] = -0.0142$$

β_1:

$$Gradiente = \frac{1}{4}[0.1874 \cdot 1 + 0.3690 \cdot 2 - 0.4028 \cdot 3 - 0.2102 \cdot 4] = -0.28095$$

Segunda Iteración

Actualización de parámetros:

- $B_0^{(2)} = -2.3968 - 0.1 \cdot (-0.0142) \approx -2.3968 + 0.0014 = -2.3954$
- $B_1^{(2)} = 0.9302 - 0.1 \cdot (-0.2809) \approx 0.9305 + 0.0281 = 0.9586$

Calculo de probabilidades

Verificamos para los valores de X, usando la ecuación:

$$p_i = \frac{1}{1 + e^{-(-2.395+0.958x_i)}}$$

Xi	**Pi(probabilidad de compra)**
1	p
2	p
3	p
4	p

Umbral de Decisión

Condición para el Umbral de Decisión en Regresión Logística:
El umbral de decisión (θ) es el valor que determina si un caso se clasifica como clase positiva (1) o negativa (0). La condición general es:

$$\text{Predicción} = \begin{cases} 0, & si\ P_i < \theta \\ 1, & si\ P_i \geq \theta \end{cases}$$

Umbral por Defecto (θ=0.5):

Condición:

Si Pi ≥0.5, se predice y=1; de lo contrario, y=0.

Xi	Pi(probabilidad de compra)
1	0.1919
2	0.3825
3	0.6175
4	0.8080

Ajuste del Umbral

Ajuste del Umbral (θ):

El umbral puede modificarse según el contexto del problema:

1. Si es crítico minimizar falsos negativos (ej: detectar fraudes):

   Se elige $\theta<0.5$ (ej: $\theta=0.3$).

2. Si es crítico minimizar falsos positivos (ej: diagnóstico médico):

   Se elige $\theta>0.5$ (ej: $\theta=0.7$).

Xi	Pi(probabilidad de compra)
1	0.1919
2	0.3825
3	0.6175
4	0.8080

Predicción

Ejemplo concreto:

Si el modelo predice que no se necesita equipo (Pi<θ), pero en realidad sí se necesita (y=1), la empresa no podrá operar!

Por lo tanto, se elige un umbral bajo θ*=0.40*

$$\text{Predicción} = \begin{cases} 0, & si\ P_i < 0.40 \\ 1, & si\ P_i \geq 0.40 \end{cases}$$

Xi	**Pi(probabilidad de compra)**	Predicción
1	0.1919	0
2	0.3825	0
3	0.6175	1
4	0.8080	1