

## LABORATORIO 05

### REGRESIÓN LOGÍSTICA

Docentes: Edward Hinojosa C. - Yesseina D. Yari R.

28 de Mayo del 2021

#### 1 CONCEPTOS BÁSICOS

- Entropía Cruzada Binaria
- Regresión Logística
- Gradiente Descendiente

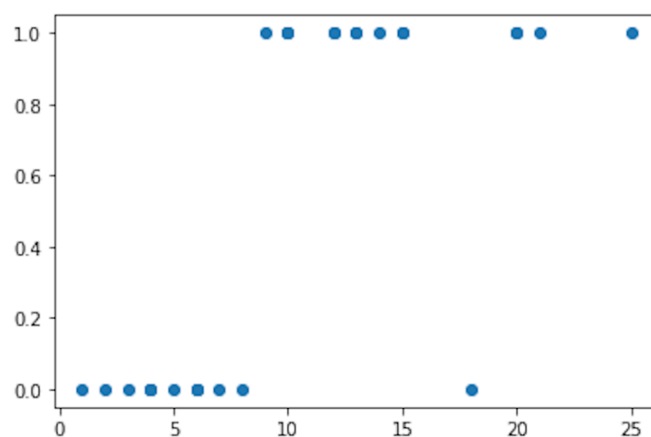
#### 2 EQUIPOS Y MATERIALES

- Un computador.
- Material del curso.
- Bibliografía del curso [1].

#### 3 EJERCICIOS

Considere los siguientes datos:

Horas Estudio	Aprobado
5	0
7	0
2	0
13	1
4	0
15	1
9	1
4	0
6	0
1	0
3	0
10	1
4	0
10	1
8	0
20	1
18	0
15	1
20	1
12	1
6	0
12	1
13	1
14	1
10	1
6	0
21	1
25	1

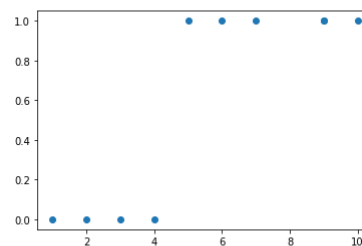


1. Seleccione 22 datos de entrenamiento (train).
2. Seleccione 6 datos de evaluación (test).
3. Implemente (en cualquier lenguaje de programación) el Método de Gradiente Descendiente para definir el modelo de Regresión Logística de los datos de entrenamiento.
  - Defina los parámetros iniciales correspondientes.
  - Genere un archivo .txt que muestre el nombre del alumno, parámetros iniciales, datos utilizados de entrenamiento y test, y por iteración: Pendiente anterior, Intercepto anterior, Error (Entropía Cruzada Binaria), Tasa de aprendizaje, Derivada pendiente, Derivada intercepto, Pendiente nueva e Intercepto nuevo.
  - Defina un umbral que debe estar en el archivo .txt.
  - Utilice el modelo de Regresión Logística para evaluar los datos de evaluación y defina la probabilidad obtenida que debe estar en el archivo .txt
  - Defina el porcentaje de acierto que debe ir en el archivo .txt

#### 4 EJEMPLO

Considere los siguientes datos:

Horas Estudio	Aprobado
5	1
7	1
2	0
9	1
3	0
10	1
9	1
4	0
6	1
1	0



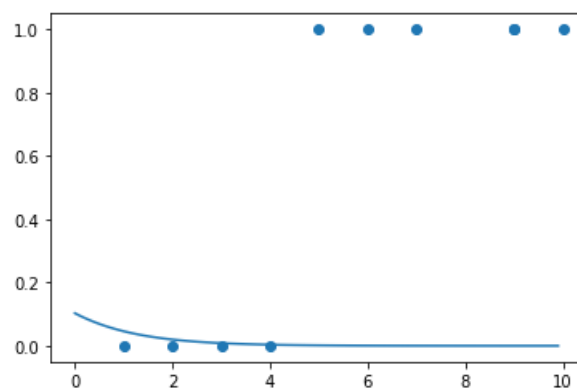
1. Seleccione 8 datos de entrenamiento (train).
2. Seleccione 2 datos de evaluación (test).
3. Defina los parámetros iniciales.
4. Implemente (en cualquier lenguaje de programación) el Método de Gradiente Descendiente para definir el modelo de Regresión Logística de los datos de entrenamiento:
  - Genere un archivo .txt que muestre el nombre del alumno, método utilizado, datos utilizados de entrenamiento y test, y por iteración: Pendiente anterior, Intercepto anterior, Error (Entropía Cruzada Binaria), Tasa de aprendizaje, Derivada pendiente, Derivada intercepto, Pendiente nueva e Intercepto nuevo.
  - Defina un umbral. Debe estar en el archivo .txt.
  - Utilice el modelo de Regresión Logística para evaluar los datos de evaluación y defina la probabilidad obtenida. Debe estar en el archivo .txt
  - Defina el porcentaje de acierto. Debe ir en el archivo .txt

Nombre del Alumno  
Pendiente anterior = -0.9874  
Intercepto anterior = -2.1789  
Tasa de aprendizaje = 0.001  
Cantidad de Iteraciones = 100000  
Umbral = 0.5  
Datos de Entrenamiento:  
 $x = \{5, 7, 2, 9, 10, 9, 4, 1\}$   
 $y = \{1, 1, 0, 1, 1, 1, 0, 0\}$   
Datos de Test:  
 $x = \{3, 6\}$   
 $y = \{0, 1\}$

Iteración 1  
Pendiente anterior = -0.9874  
Intercepto anterior = -2.1789  
Error = 50.450518  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -39.914733  
Derivada intercepto = -4.940947  
Pendiente nueva = -0.9475  
Intercepto nuevo = -2.174

Iteración 2  
Pendiente anterior = -0.947485  
Intercepto anterior = -2.173959  
Error = 48.833072  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -39.907381  
Derivada intercepto = -4.937198  
Pendiente nueva = -0.9076  
Intercepto nuevo = -2.169

Iteración 3  
Pendiente anterior = -0.907578  
Intercepto anterior = -2.169022  
Error = 47.216269  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -39.899088  
Derivada intercepto = -4.933131  
Pendiente nueva = -0.8677  
Intercepto nuevo = -2.1641

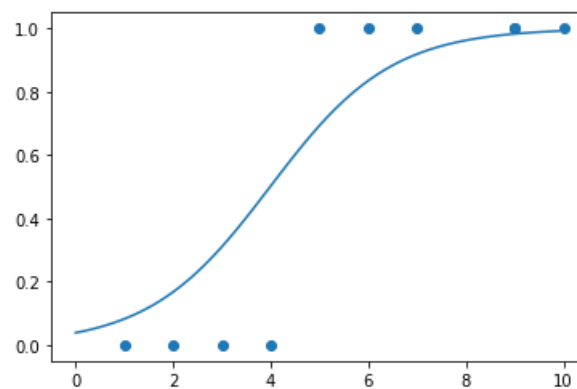


Iteración 2962  
Pendiente anterior = 0.805765  
Intercepto anterior = -3.213166  
Error = 1.461604  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -0.063047  
Derivada intercepto = 0.322588  
Pendiente nueva = 0.8058  
Intercepto nuevo = -3.2135

Iteración 2963  
Pendiente anterior = 0.805828  
Intercepto anterior = -3.213489  
Error = 1.461496  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -0.063039  
Derivada intercepto = 0.322542  
Pendiente nueva = 0.8059  
Intercepto nuevo = -3.2138

Iteración 2964  
Pendiente anterior = 0.805891  
Intercepto anterior = -3.213811  
Error = 1.461388  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -0.063031  
Derivada intercepto = 0.322495  
Pendiente nueva = 0.806  
Intercepto nuevo = -3.2141

Iteración 2965  
Pendiente anterior = 0.805954  
Intercepto anterior = -3.214134  
Error = 1.46128  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -0.063023  
Derivada intercepto = 0.322448  
Pendiente nueva = 0.806  
Intercepto nuevo = -3.2145



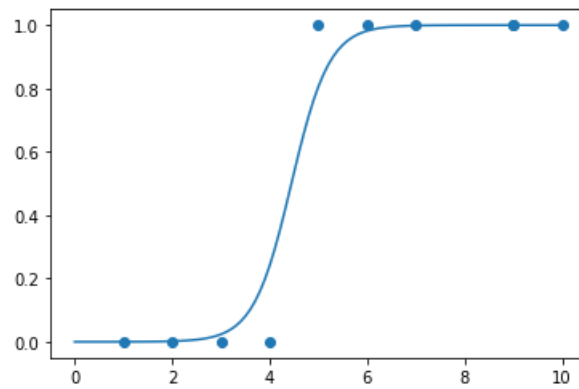
Iteración 99998  
Pendiente anterior = 2.574016  
Intercepto anterior = -11.443609  
Error = 0.494303  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -0.010581  
Derivada intercepto = 0.047859  
Pendiente nueva = 2.574  
Intercepto nuevo = -11.4437

Iteración 99999  
Pendiente anterior = 2.574026  
Intercepto anterior = -11.443657  
Error = 0.4943  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -0.010581  
Derivada intercepto = 0.047859  
Pendiente nueva = 2.574  
Intercepto nuevo = -11.4437

Iteración 100000  
Pendiente anterior = 2.574037  
Intercepto anterior = -11.443705  
Error = 0.494298  
Tasa de aprendizaje = 0.001  
Derivada pendiente = -0.010581  
Derivada intercepto = 0.047859  
Pendiente nueva = 2.574  
Intercepto nuevo = -11.4438

#### Test

Primer dato 3 = 0.02362340010368705, aprobado estimado = 0, Correcto  
Primer dato 6 = 0.9820231806942471, aprobado estimado = 1, Correcto  
Porcentaje de Acierto: 100.00%



## 5 ENTREGABLES

Al finalizar el estudiante deberá:

1. Generar un gráfico con el resultado obtenido (Regresión Logística) con todos los datos.
2. Generar el archivo .txt solicitado.
3. Comprimir en un archivo .zip todos los archivos anteriores (además de todo código fuente, sin el cual no se revisará el laboratorio) y subirlo el archivo al aula virtual hasta 5 minutos después de finalizar la hora del laboratorio con el nombre:

Laboratorio\_XX\_ApellidoPaterno\_ApellidoMaterno\_PrimerNombre\_IA\_2021A\_EPCC\_UNSA.zip

## 6 RÚBRICA DE EVALUACIÓN

Criterios	Muy Bueno	Bueno	Regular	Malo
Resolución del Laboratorio	Resuelve los ejercicios sin errores mostrando cada uno de los puntos solicitados. <b>Puntaje: 16-20 puntos</b>	Resuelve los ejercicios con pocos errores mostrando casi o todos todos los puntos solicitados. <b>Puntaje: 11-15 puntos</b>	Resuelve los ejercicio con varios errores y mostrando todos o pocos de los puntos solicitados. <b>Puntaje: 1-10 puntos</b>	No resuelve los ejercicios o no entrega el laboratorio. <b>Puntaje: 0 puntos</b>

- **IMPORTANTE** En caso de copia o plagio o similares todos los alumnos implicados tendrán sanción en toda la evaluación del curso.

## BIBLIOGRAFÍA

- [1] SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Database System Concepts**. 7th. Edition: McGraw-Hill Book Company, 2020. ISBN 9780078022159.