



Universidad Nacional Autónoma de México
Facultad de Ingeniería

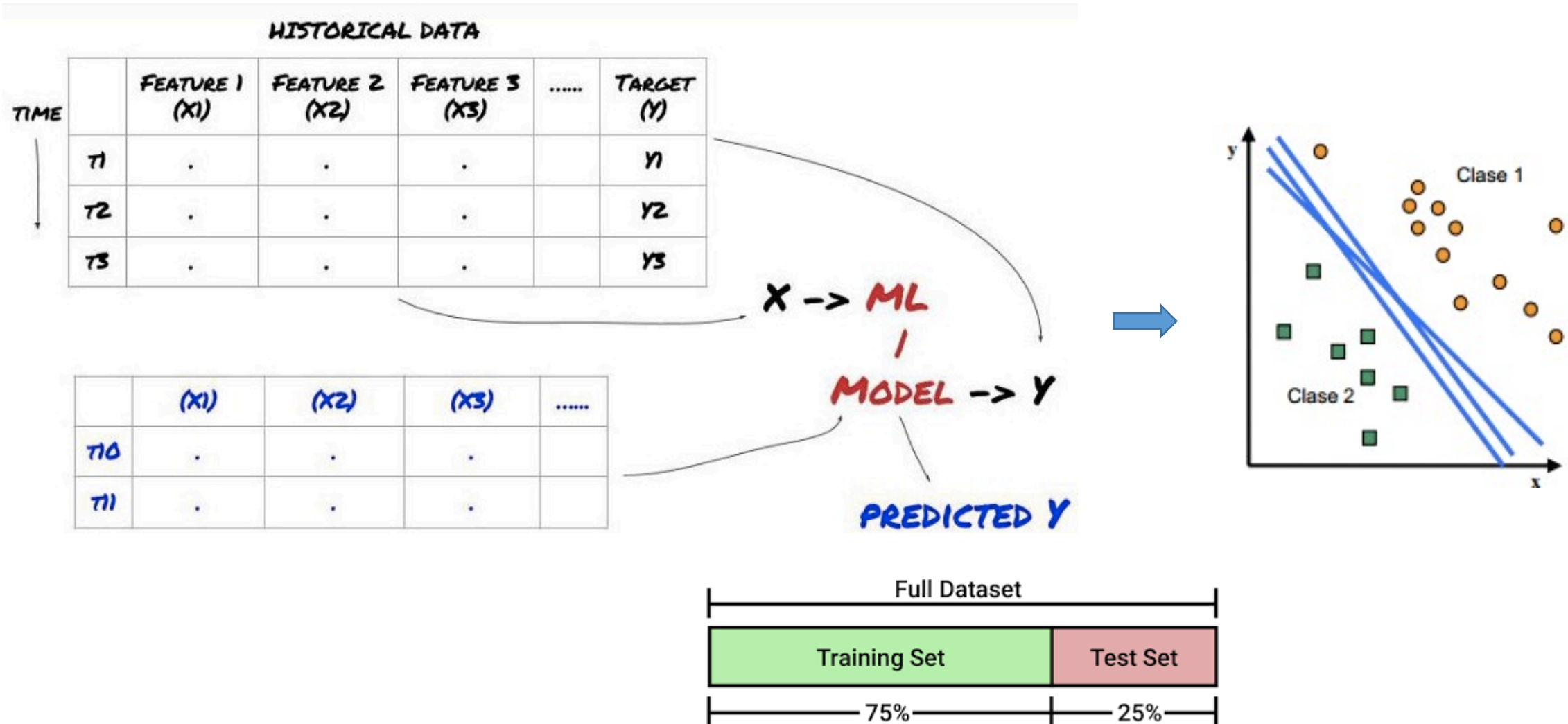
Regresión Logística

Práctica

Clasificación

Enero, 2021

Contexto



Caso práctico

Caso de estudio

Registros clínicos de cáncer de mama a partir de imágenes digitalizadas.
La *Tabla* muestra las variables oncológicas disponibles.

Variable	Descripción	Tipo
ID number	Identifica al paciente	Discreto
Diagnosis	Diagnostico (M=maligno, B=benigno)	Booleano
Radius	Media de las distancias del centro y puntos del perímetro	Continuo
Texture	Desviación estándar de la escala de grises	Continuo
Perimeter	Valor del perímetro del cáncer de mama	Continuo
Area	Valor del área del cáncer de mama	Continuo
Smoothness	Variación de la longitud del radio	Continuo
Compactness	$\text{Perímetro}^2 / \text{Area} - 1$	Continuo
Concavity	Caída o gravedad de las curvas de nivel	Continuo
Concave points	Número de sectores de contorno cóncavo	Continuo
Symmetry	Simetría de la imagen	Continuo
Fractal dimension	Aproximación de frontera - 1	Continuo

Caso de estudio

Registros clínicos de cáncer de mama a partir de imágenes digitalizadas.

Identificador	Diagnosis	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	Concave points	Symmetry	Fractal dimension
P-842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871
P-842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667
P-84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999
P-84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744
P-84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883
P-843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613
P-844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742
P-84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451
P-844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389
P-84501001	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243

M = Maligno

B = Benigno

Caso de estudio

- Registros de imágenes digitalizadas de **569** pacientes
- Variables independientes 10
- Variable dependiente 1 (**Diagnóstico**)
- Omitir la variable **Identificador**.

1	Identificador	Diagnosis	Radius	Texture	Perimeter	Area	Smoothness	Compactness	Concavity	Concave points	Symmetry	Fractal dimension
2	P-842302	0	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871
3	P-842517	0	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667
4	P-84300903	0	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999
5	P-84348301	0	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744
6	P-84358402	0	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883
7	P-843786	0	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613
8	P-844359	0	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742
9	P-84458202	0	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451
10	P-844981	0	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389
11	P-84501001	0	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243
12	P-845636	0	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528	0.05697

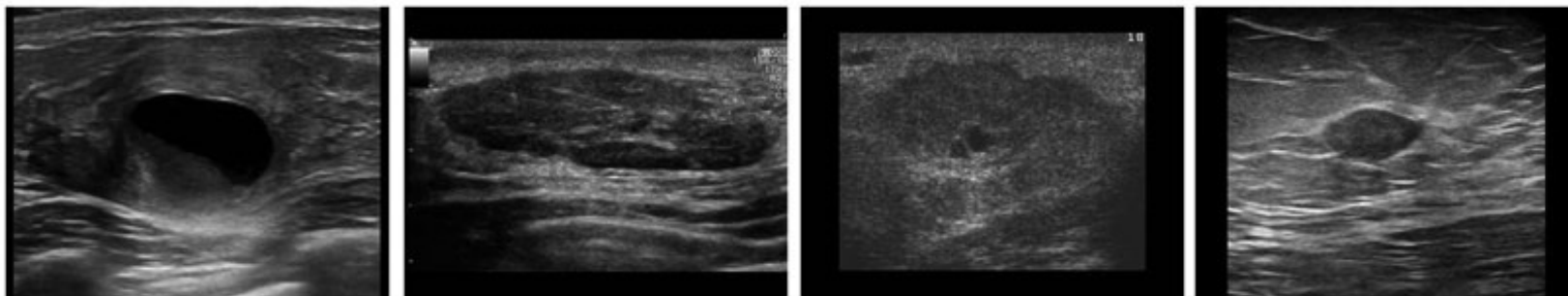
0 = Maligno; 1 = Benigno

De las 569 observaciones, 357 son tumores benignos y 212 malignos

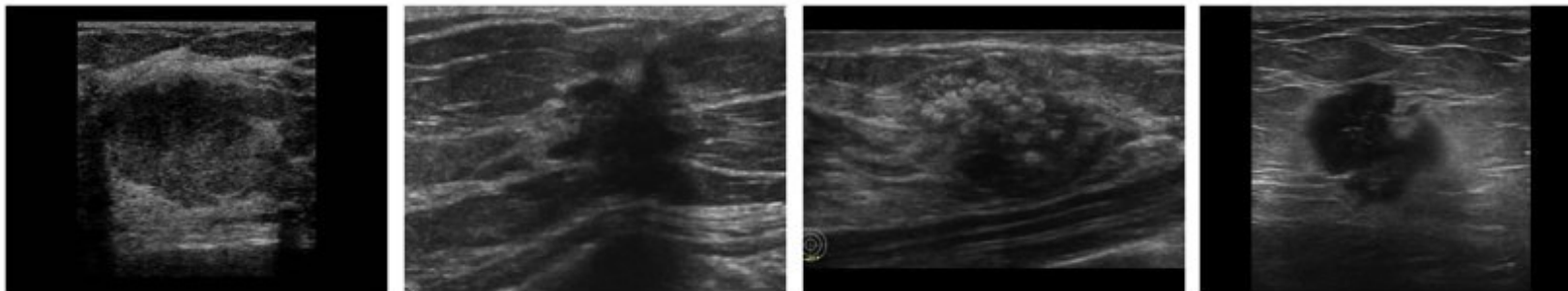
Caso de estudio

Registros clínicos de cáncer de mama a partir de imágenes digitalizadas.

Benigno

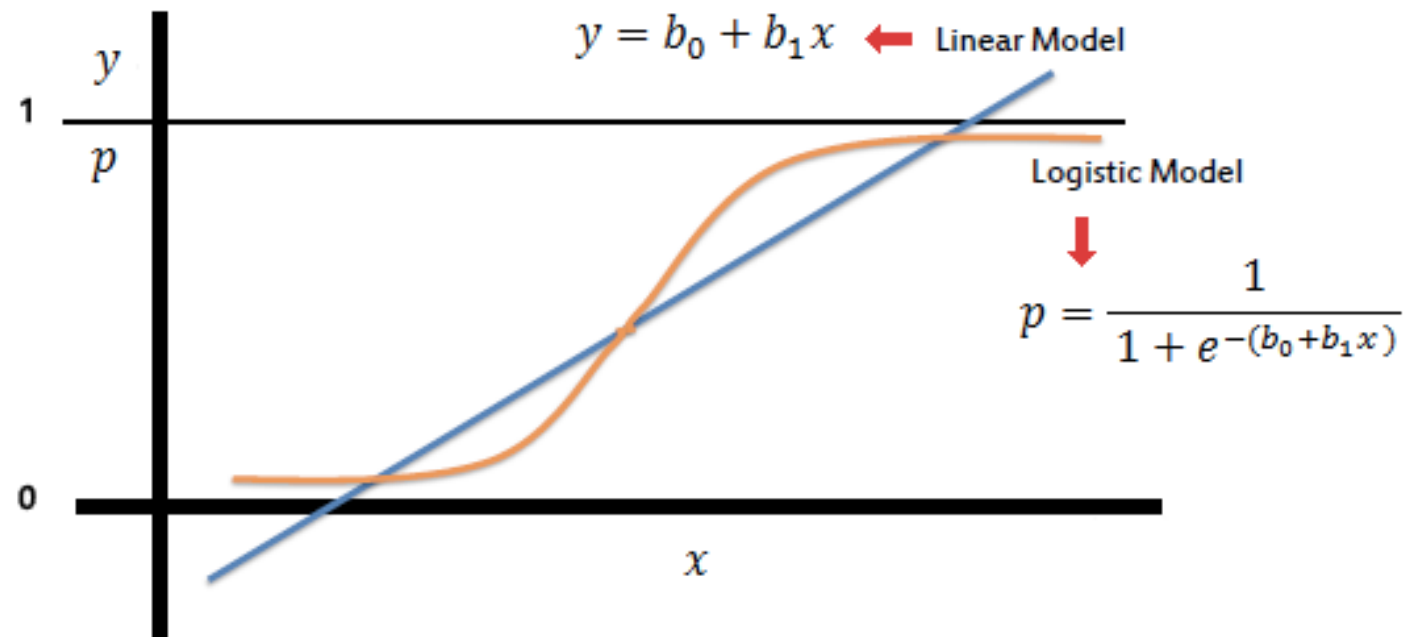


Maligno



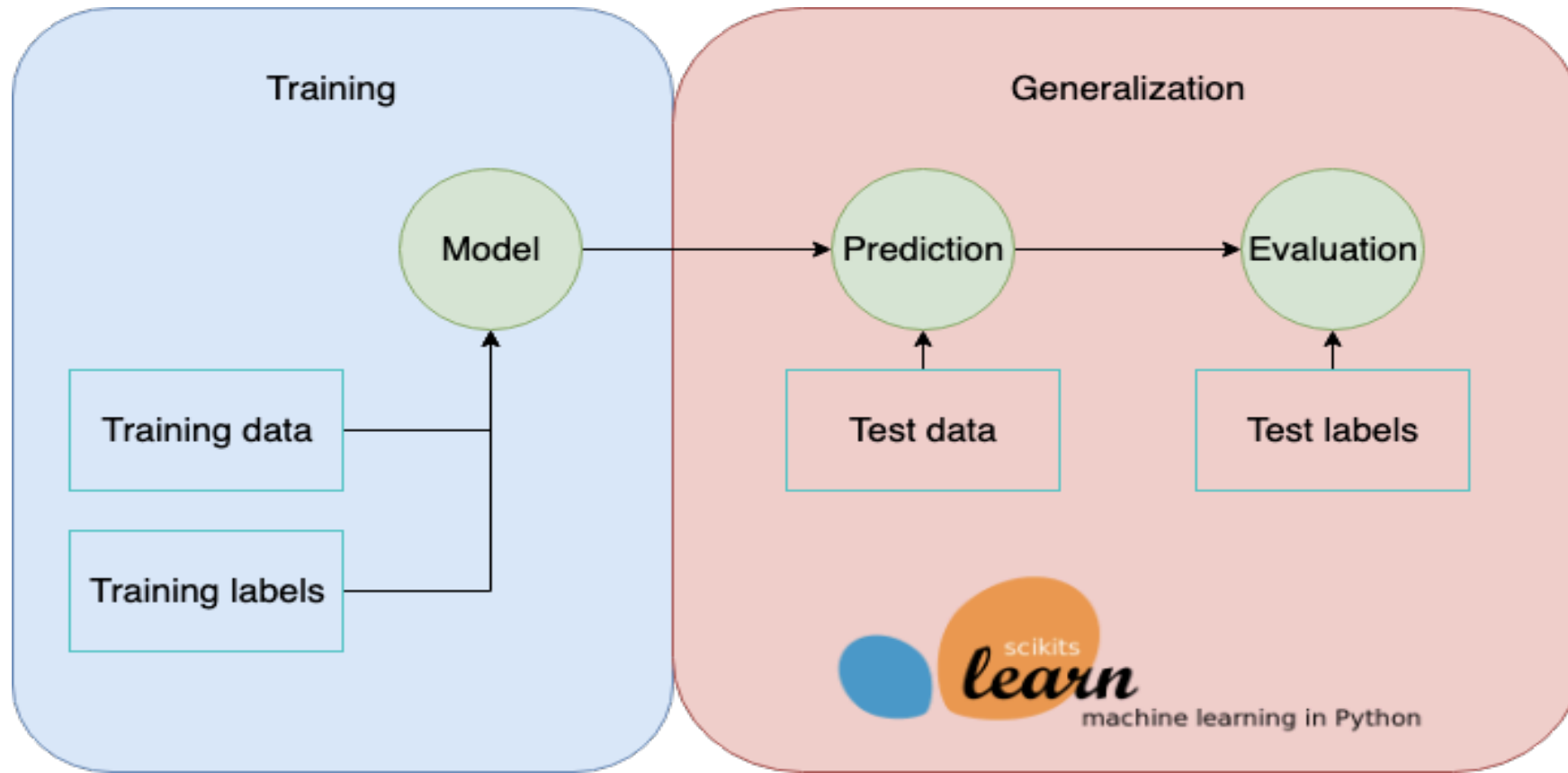
Regresión logística

- Diagnóstico (variable dependiente)



Google Colab

Regresión logística



Matriz de clasificación

Predicción	0	1
Real		
0	39	6
1	2	67

Exactitud 0.9298245614035088				
	precision	recall	f1-score	support
0	0.95	0.87	0.91	45
1	0.92	0.97	0.94	69
accuracy			0.93	114
macro avg	0.93	0.92	0.93	114
weighted avg	0.93	0.93	0.93	114

$Exactitud = \frac{VP+VN}{VP+VN+FP+FN} = \frac{67+39}{167+39+2+614} = \frac{106}{114} = 0.929$

$Precisión = \frac{VP}{VP+FP} = \frac{67}{67+6} = \frac{67}{73} = 0.917$

$Tasa\ de\ error = \frac{FP+FN}{VP+VN+FP+FN} = \frac{2+6}{167+39+2+614} = \frac{8}{114} = 0.07$

$Especificidad = \frac{VN}{VN+FP} = \frac{39}{39+6} = \frac{39}{45} = 0.866$

$Sensibilidad = \frac{VP}{VP+FN} = \frac{67}{67+2} = \frac{67}{69} = 0.971$

Modelo de clasificación

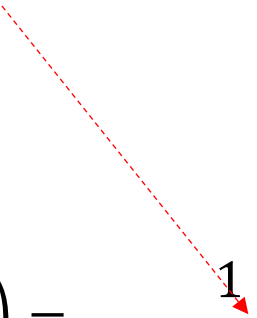
```
[ ] #Información del modelo
print("Intercept:", Clasificacion.intercept_)
print('Coeficientes: \n', Clasificacion.coef_)
```

Intercept: [11.72346938]

Coeficientes:

[[-0.1908854 -0.0106113 -2.27353782 -3.0783869 -0.87661688 -0.2123106]]

a+bX = 11.72 - 0.19Texture - 0.01Area - 2.27Compactness - 3.08Concavity - 0.88Symmetry - 0.21FractalDimension

$$\ln\left(\frac{P}{1-P}\right) = \frac{1}{1 + e^{-(a+bX)}}$$


Sistema de inferencia basado en un modelo de regresión logística

ID Paciente	<input type="text"/>		
Texture	<input type="text"/>	Concavity	<input type="text"/>
Area	<input type="text"/>	Symmetry	<input type="text"/>
Compactness	<input type="text"/>	Fractal Dimension	<input type="text"/>
Diagnosis		<input type="text"/>	