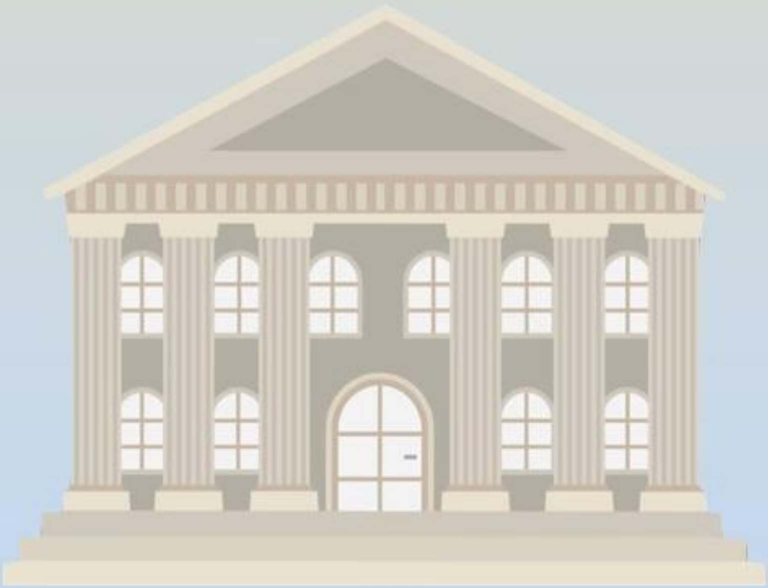
Regresión Logística

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ingreso a Universidad

Examen

Ingreso a Universidad



Examen

Notas Secundario

Ingreso a Universidad

Examen **Notas Secundario**

Estudiante 1:
Examen: 9
Notas Sec: 8

Ingreso a Universidad

Examen **Notas Secundario**

Estudiante 1:
Examen: 9
Notas Sec: 8

Estudiante 2 :
Examen: 3
Notas Sec: 4

Ingreso a Universidad

Examen **Notas Secundario**

Estudiante 1:
Examen: 9
Notas Sec: 8

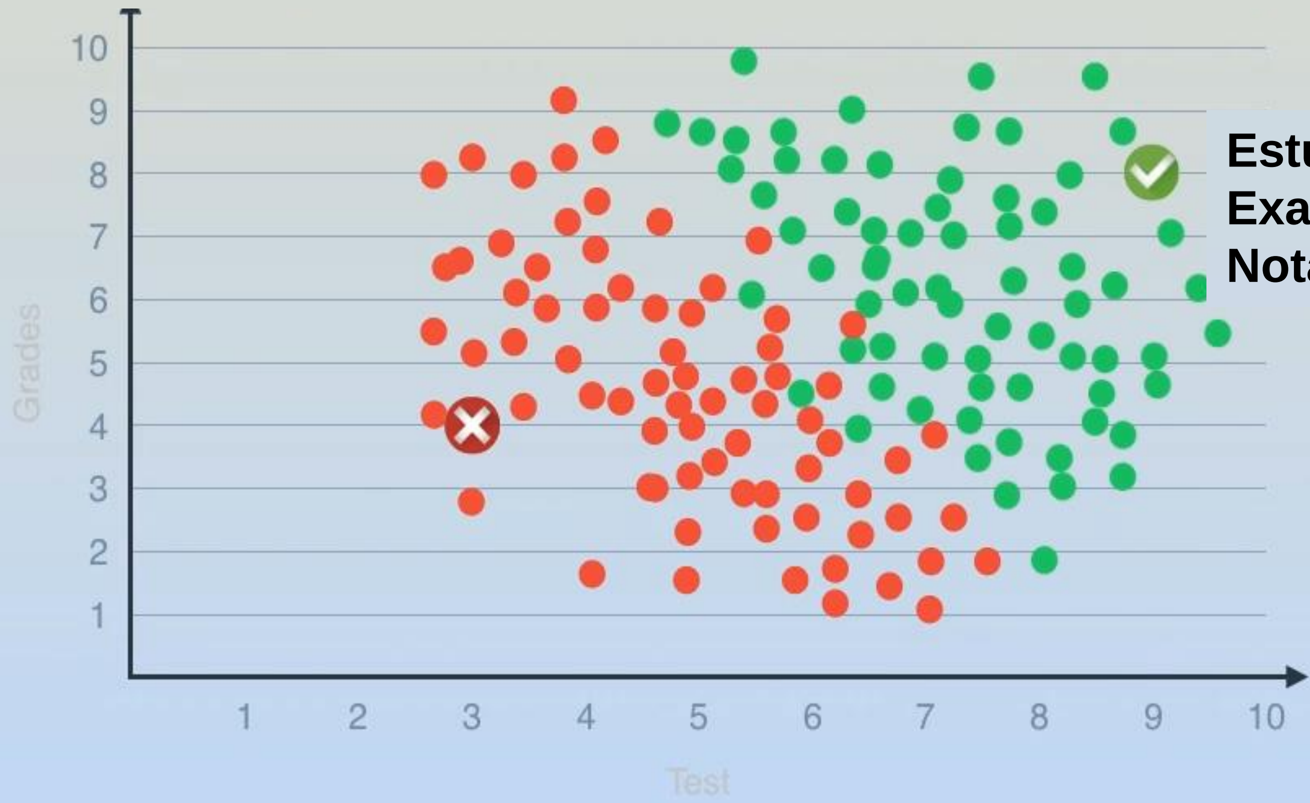
Estudiante 2 :
Examen: 3
Notas Sec: 4

Estudiante 3 :
Examen: 7
Notas Sec: 6

Ingreso a Universidad

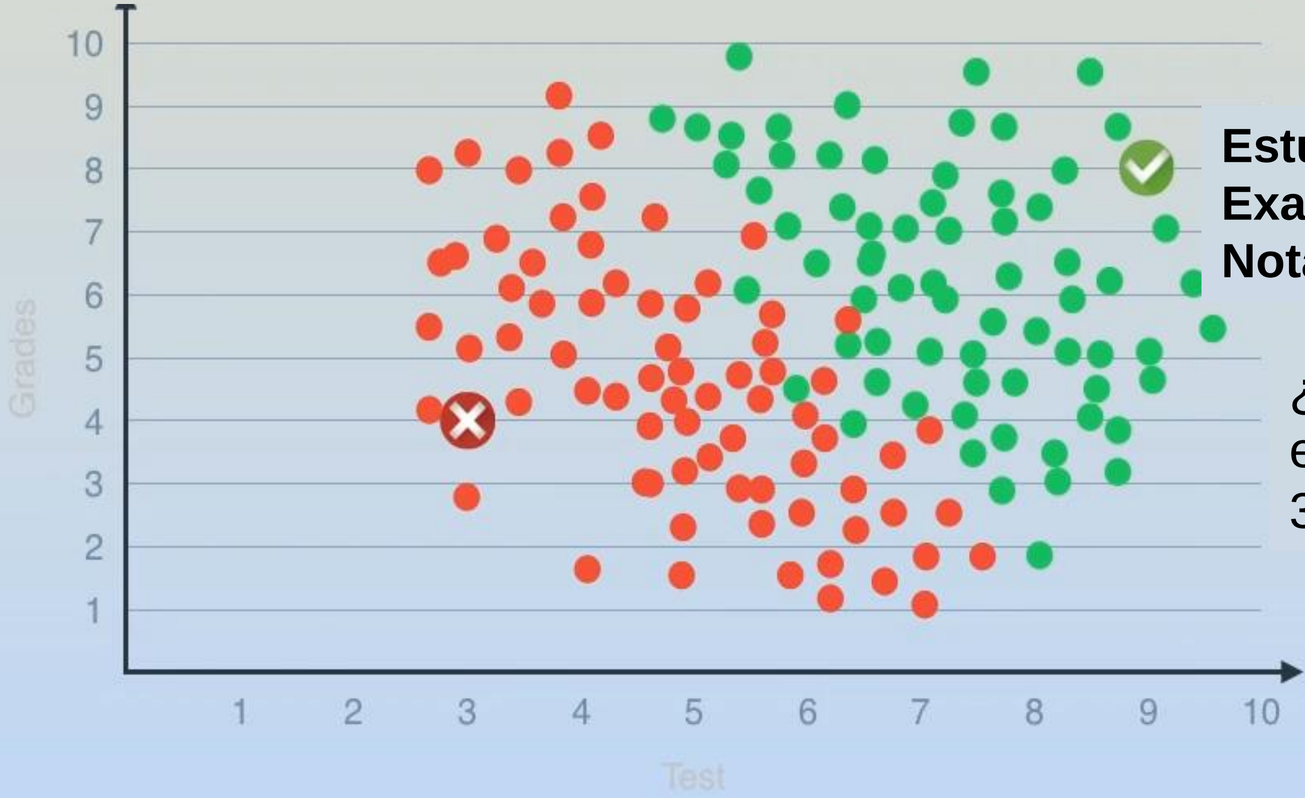
Estudiante 3 :
Examen: 7
Notas Sec: 6

Ingreso a Universidad



Estudiante 3 :
Examen: 7
Notas Sec: 6

Ingreso a Universidad

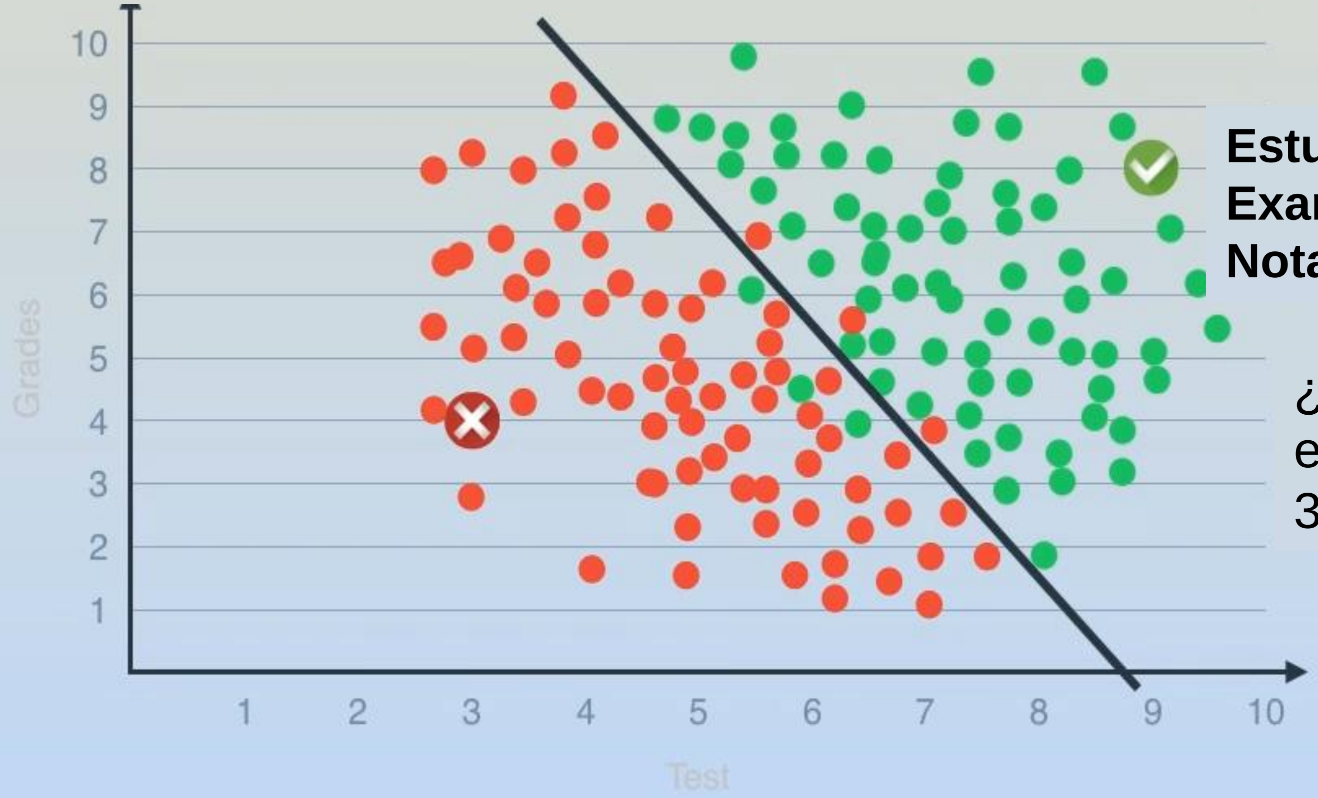


Estudiante 3 :
Examen: 7
Notas Sec: 6

¿Es aceptado el Estudiante 3 o no ?

○ **SI**

Ingreso a Universidad

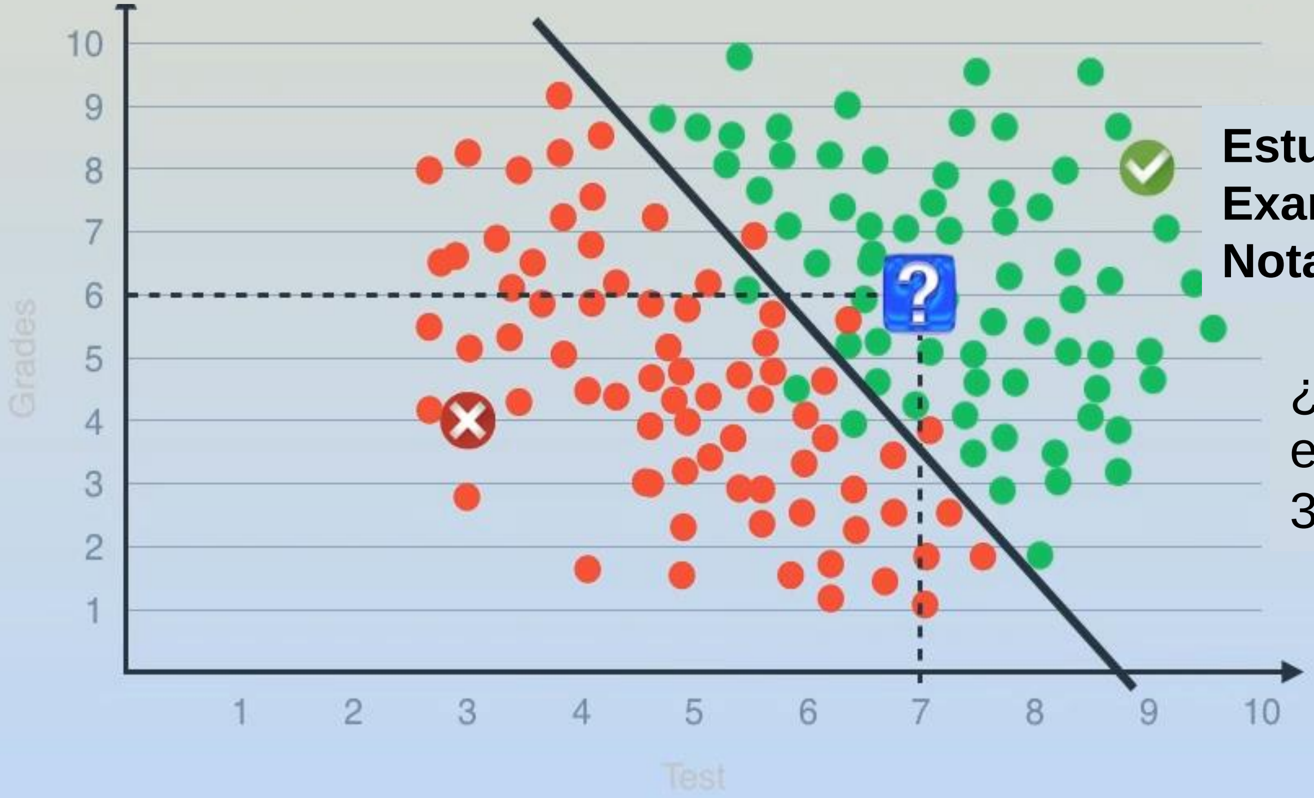


Estudiante 3 :
Examen: 7
Notas Sec: 6

¿Es aceptado el Estudiante 3 o no ?

SI

Ingreso a Universidad



Estudiante 3 :
Examen: 7
Notas Sec: 6

¿Es aceptado el Estudiante 3 o no ?

○ **SI**

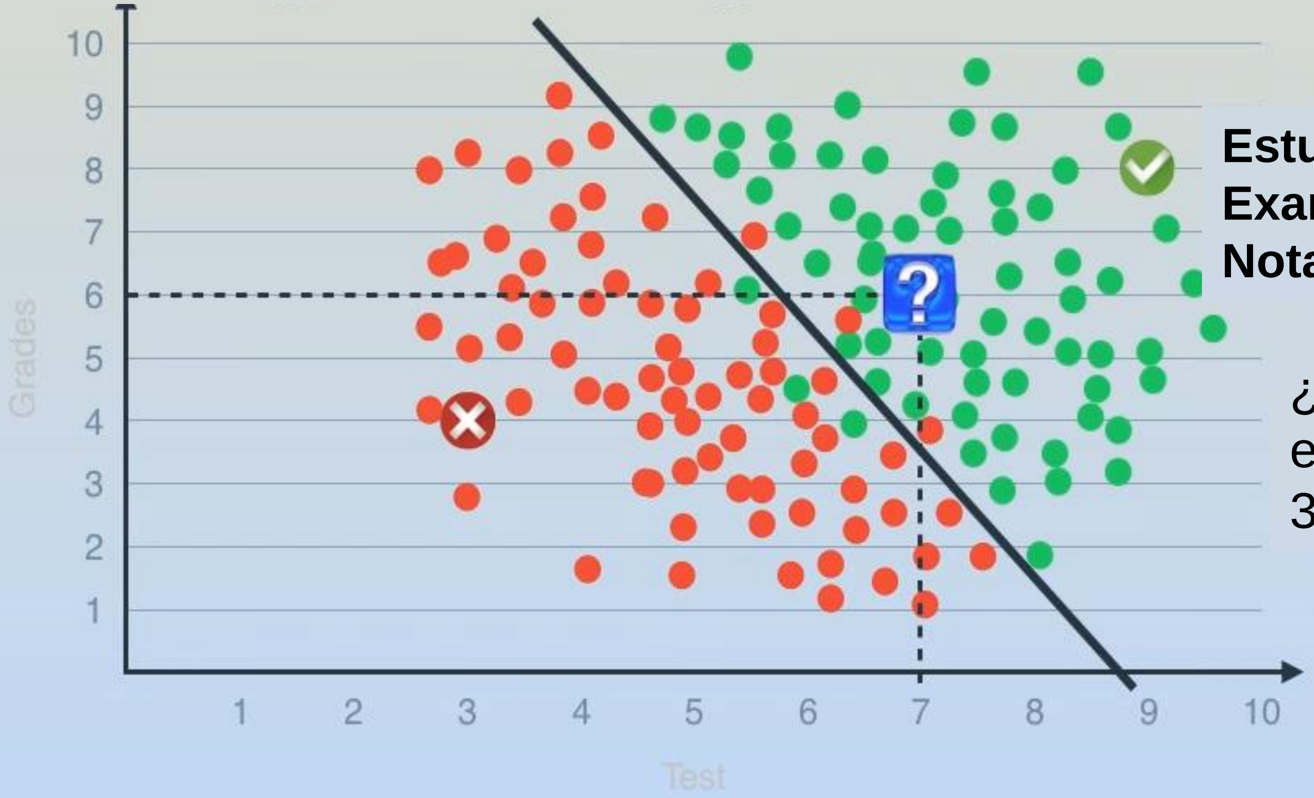
Ingreso a Universidad

Estudiante 3 :
Examen: 7
Notas Sec: 6

¿Es aceptado el Estudiante 3 o no ?

SI

Regresión Logística

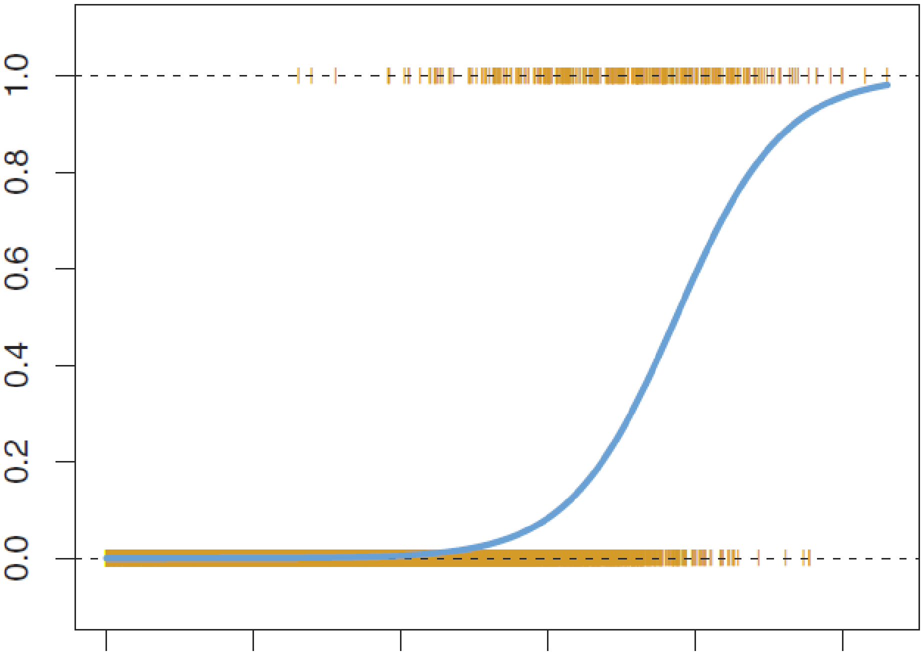


Estudiante 3 :
Examen: 7
Notas Sec: 6

¿Es aceptado el Estudiante 3 o no ?

SI

¿Cómo funciona La Regresión Logística?



$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

El modelo Logístico

La regresión logística modela la **probabilidad** que la variable respuesta tome un cierto valor dadas las *p* variables predictoras. Ejemplo: expresaremos la **probabilidad** de que la variable categórica RESPUESTA 'default' sea POSITIVA(en este caso que tome el valor **"yes")** dado un cierto valor observado de la variable cuantitativa EXPLICATIVA (o predictora) **'balance'** como:

$$\Pr(\texttt{default} = \texttt{Yes}|\texttt{balance}).$$

Esta probabilidad, por razones obvias, debe estar comprendida entre 0 y 1. **A partir de ella deberemos tomar una decisión para clasificar a la variable respuesta en una de las categorías posibles**, en este caso 1, YES, o TRUE para indicar *Default*, o 0, NO o FALSE para indicar lo contrario. Lo más usual, que es la denominada NEUTRAL, sería:

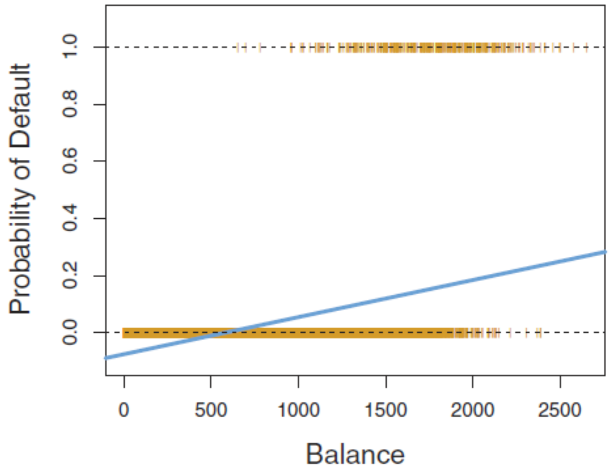
$$\Pr(\texttt{default} = \texttt{Yes}|\texttt{balance}) > 0.5 \quad \Rightarrow \quad \texttt{default} = \texttt{Yes}$$

Este límite de 0,5 depende del entorno en el que se haga la predicción. Podría ser el caso que la institución o empresa tenga una política de selección muy **conservadora** por lo que sería muy estricta al clasificar como Default , entonces podría seguir la sig regla de clasificación:

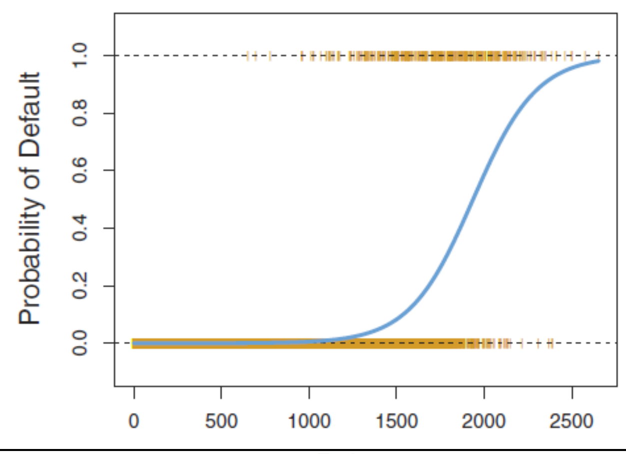
$$\Pr(\texttt{default} = \texttt{Yes}|\texttt{balance}) > 0.1 \quad \Rightarrow \quad \texttt{default} = \texttt{Yes}$$

¿Qué pasaría si utilizáramos el modelo de regresión lineal para predecir la variable *Default*?

$$p(X) = \beta_0 + \beta_1 X$$



A todas luces este modelo no es el apropiado pues estaría indicando probabilidades negativas para ciertos valores de la variable predictora. Por ende deberíamos utilizar modelo cuya variable dependiente tome valores entre **0 y 1**, y así asimilarla a una **probabilidad.**



Probabilidades vs Chance

- La **probabilidad** de que ocurra un evento es la fracción de veces que se espera que se ocurra ese evento en muchos ensayos o experimentos. Las probabilidades siempre toman valores entre 0 y 1.
- Las **chances** se definen como la probabilidad de que el **evento ocurra** dividida entre la probabilidad de que **el evento no ocurra**.

Una probabilidad de 0,5 es la misma que las probabilidades de 1,0.

Ejemplo: La probabilidad de lanzar una moneda y que salga cara es del 50%. Las chances son "cincuenta: cincuenta", lo que equivale a 1,0.A medida que la probabilidad sube de 0.5 a 1, las chances aumentan de 1 a casi infinito. Por ejemplo, si la probabilidad es 0.75, entonces las chances son 75:25, tres a uno, o 3.0.Si las chances son altas (un millón a uno), la probabilidad es de casi 1.00. Si las probabilidades son minúsculas (de uno a un millón), la probabilidad es minúscula, casi nula.

Convertir entre probabilidades y probabilidad es sencillo:

$$\text{Chances (x)} = \frac{p(x)}{1-p(x)}$$

Se pude demostrar matemáticamente de forma sencilla que el logaritmo natural de una chance sí tiene un dependencia lineal con ***x***, al estilo de una regresión lineal simple.

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$$

Sí es lineal !!!

Aplicando logaritmos a ambos lados queda:

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$$

Y despejando obtrenemos la denominada Función logística o sigmoidea:

$$\boxed{p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}$$

o

$$\boxed{p(X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}}}$$

Estimación de los coeficientes

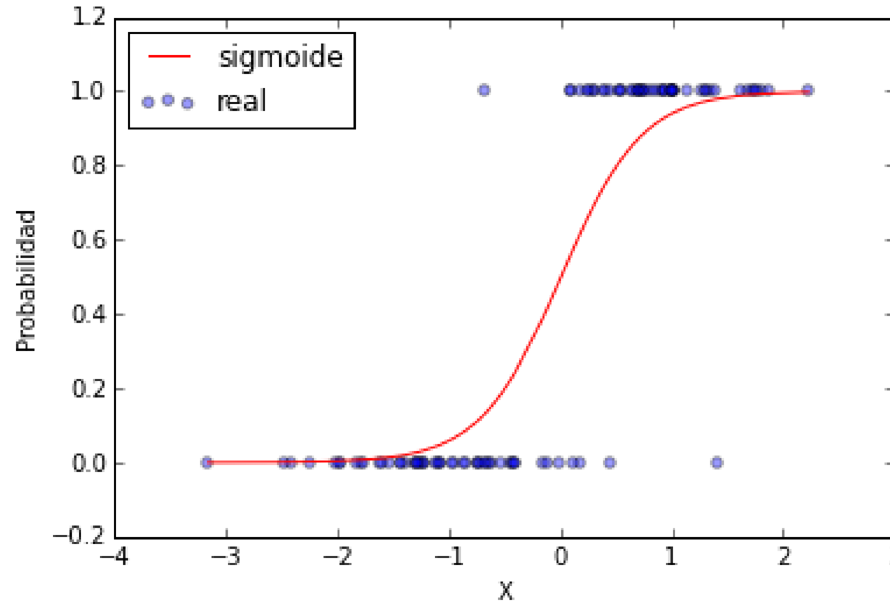
$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$

$$\begin{aligned} P(y=1 \mid x;\theta) &= h_\theta(x) \\ P(y=0 \mid x;\theta) &= 1 - h_\theta(x) \end{aligned}$$

$$p(y \mid x;\theta) = (h_\theta(x))^y \, (1 - h_\theta(x))^{1-y}$$

Si tengo m observaciones en la muestra (el dateset) , ellas podrán tener valor $y_i = 1$ o $y_i=0$

$$\begin{aligned} L(\theta) &= p(\vec{y} \mid X;\theta) \\ &= \prod_{i=1}^{m} p(y^{(i)} \mid x^{(i)};\theta) \\ &= \prod_{i=1}^{m} \left(h_\theta(x^{(i)})\right)^{y^{(i)}} \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}} \end{aligned}$$



Maximizar la función L es lo mismo que maximizar su logaritmo, y éste último es más fácil de derivar

$$\begin{aligned} \ell(\theta) &= \log L(\theta) \\ &= \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)})) \end{aligned}$$

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

$$p(X) = \frac{e^{\beta_0+\beta_1X_1+\cdots\ldots+\beta_pX_p}}{1+e^{\beta_0+\beta_1X_1+\cdots\ldots+\beta_pX_p}}$$

$$p(X) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\cdots\ldots+\beta_pX_p)}}$$