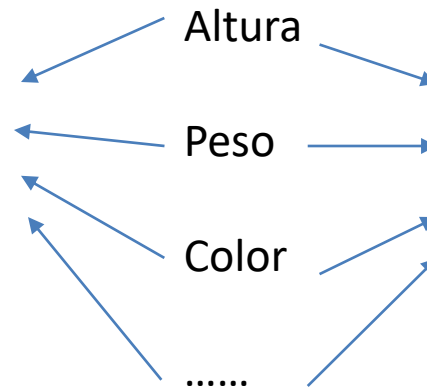




Características (features)

- Son las propiedades medibles de un objeto observado.
- Pueden ser independientes (predictoras) o dependientes (predichas)
- Pueden ser numéricas (reales o enteras) o no (texto, grafos,..)





dataset

- Las observaciones sobre los objetos se suelen representar en forma matricial en datasets. Cada fila representa una observación y cada columna una característica medible del objeto

Iris dataset

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Features
(attributes, measurements, dimensions)

Class labels
(targets)

Petal

Sepal



Problemas

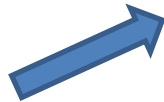
- Predecir (clasificación)

Altura

Peso

Color

.....



A partir de estas características... es un elefante o un ratón ?



Problemas

- Predecir (regresión)

Nº de baños

Distrito

Ciudad

Superficie

Nº de terrazas

.....

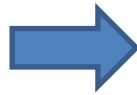


A partir de estas características... cual será su precio de venta?

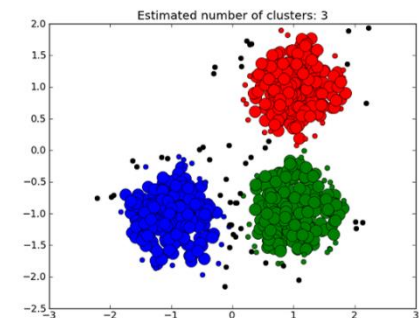


Problemas

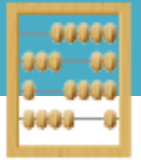
- Descubrir (clustering)



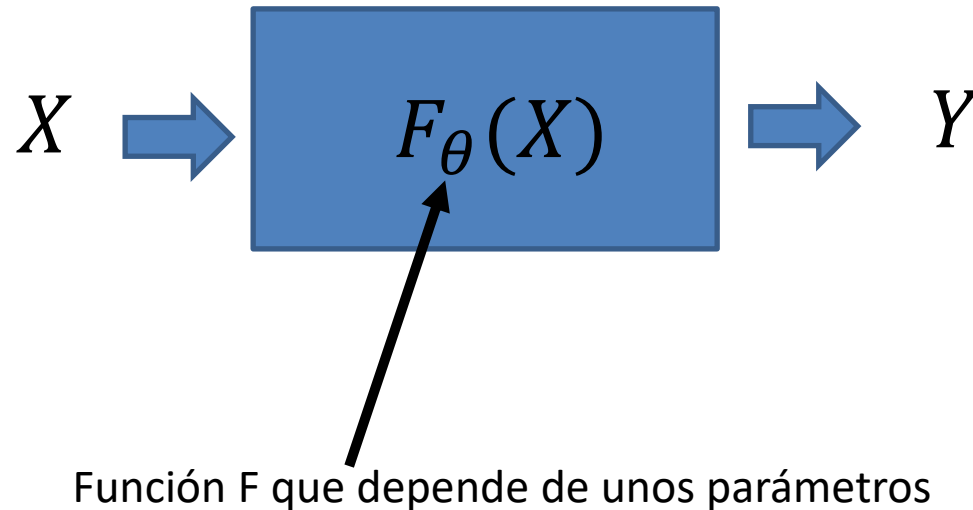
Nº de baños
Distrito
Ciudad
Superficie
Nº de terrazas
.....



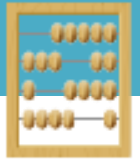
A partir de estas características de muchas casas, podemos asignar automáticamente las casas a categorías sin conocer estas categorías previamente?



Aprendizaje supervisado

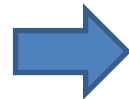
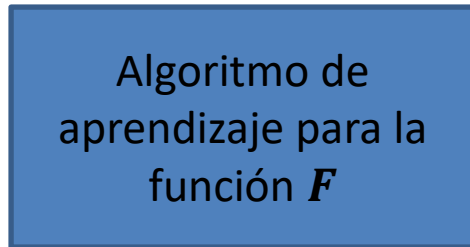


Los parámetros θ los **aprendemos** de manera **automática** a partir de datos de entrenamiento X_{train} , Y_{train}



Aprendizaje supervisado

X_{train}, Y_{train}

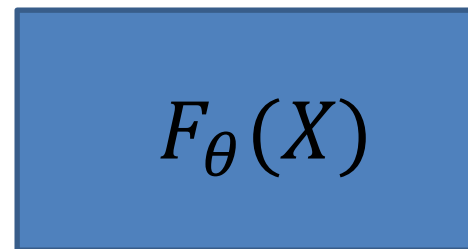


Aprendemos θ



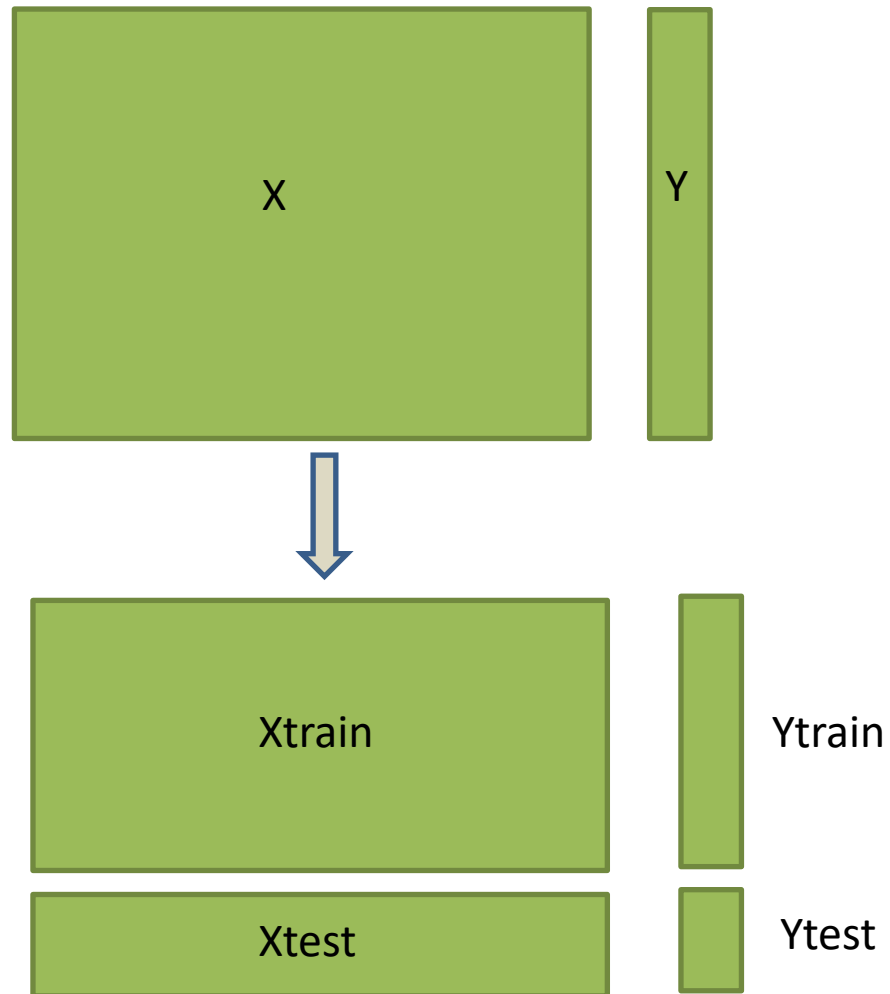
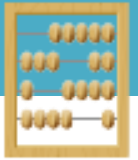
Usamos el θ aprendido

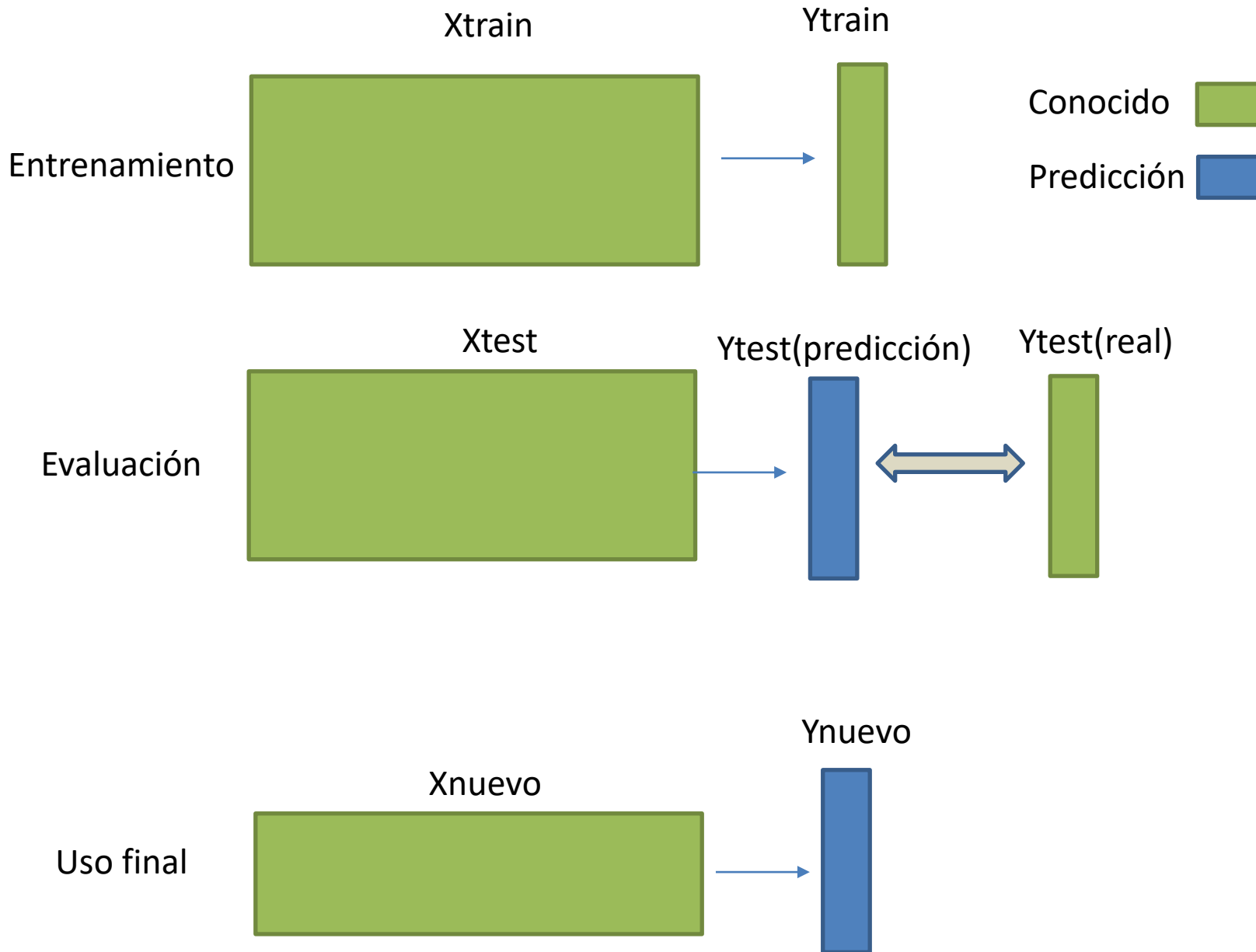
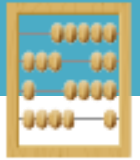
X_{test}

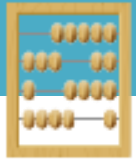


Y_{test}

Usamos la función genérica F con el parámetro θ **ajustado a los datos de entrenamiento** para predecir Y para unos datos X de entrada **desconocidos**







Aprendizaje supervisado → Ejemplo , regresión

$$F_{\theta}(X) \Rightarrow \hat{Y} = a X_1 + b X_2 + c X_3 + d \quad \leftarrow \text{Regresión lineal}$$

$$\theta \Rightarrow [a, b, c, d]$$

Algoritmo de aprendizaje



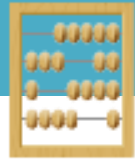
Minimización de la función de coste:

$$X_{train}: X_{1i}, X_{2i}, X_{3i}$$

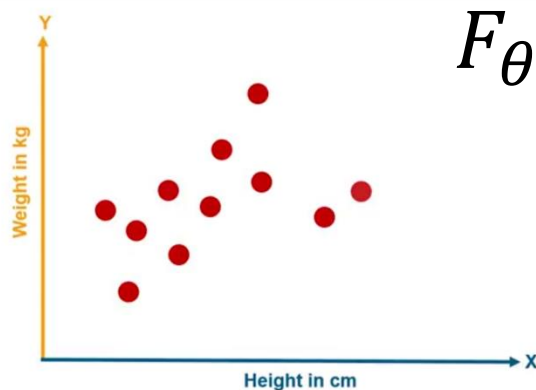
$$Y_{train}: Y_i$$

$$\underset{a,b,c,d}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^N [Y_i - (a X_{1i} + b X_{2i} + c X_{3i} + d)]^2$$

- Ecuaciones normales (Normal equations)
- Descenso del gradiente (Gradient Descent)
- Métodos bayesianos



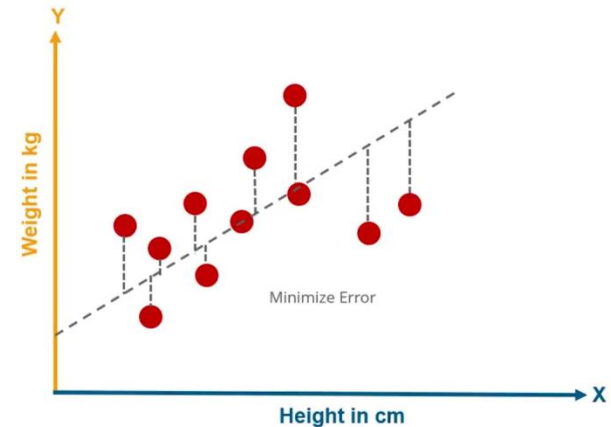
Aprendizaje supervisado → Ejemplo, regresión



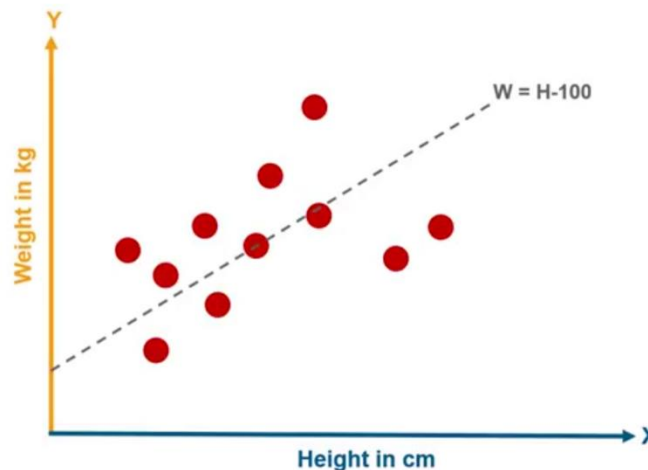
$$F_{\theta}(X) \Rightarrow Y = a X_1 + b$$



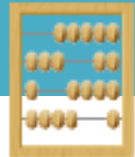
Modelo



Optimización



$$\operatorname{argmin}_{a,b} \frac{1}{N} \sum_{i=1}^N [Y_i - (a X_{1i} + b)]^2$$



Aprendizaje supervisado → Ejemplo, clasificación

$$F_{\theta}(X) \Rightarrow \hat{Y}(X, \theta) = \frac{1}{1 + e^{-(aX_1 + bX_2 + cX_3 + d)}} \quad \leftarrow \text{Regresión logística}$$

$$\theta \Rightarrow [a, b, c, d]$$

Se asocia a la probabilidad de tener un valor positivo

Algoritmo de aprendizaje



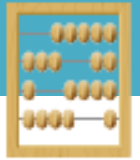
Minimización de la función de coste:

$X_{train}: X_{1i}, X_{2i}, X_{3i}$

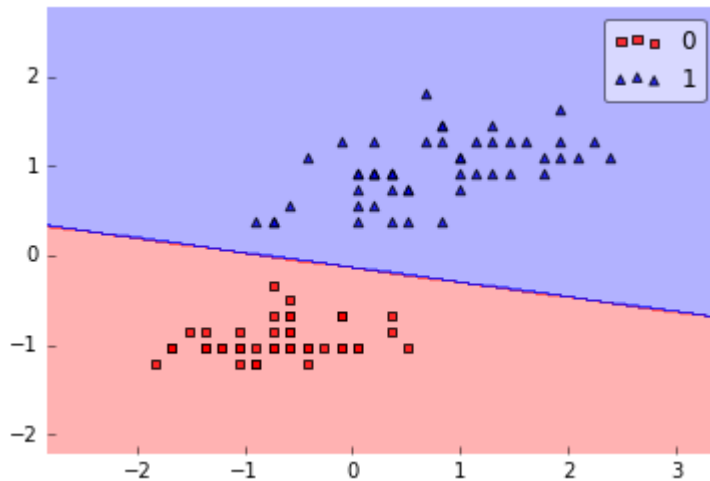
$Y_{train}: Y_i$

$$\operatorname{argmin}_{a,b,c,d} -\frac{1}{N} \sum_{i=1}^N [Y_i \log(\hat{Y}(X_i, \theta)) + (1 - Y_i) \log(1 - \hat{Y}(X_i, \theta))]$$

- Descenso del gradiente (Gradient Descent)



Aprendizaje supervisado → Ejemplo, clasificación



Modelo $F_{\theta}(X)$

$$\hat{Y}(X, \theta) = \frac{1}{1 + e^{-(aX_1 + bX_2 + c)}}$$

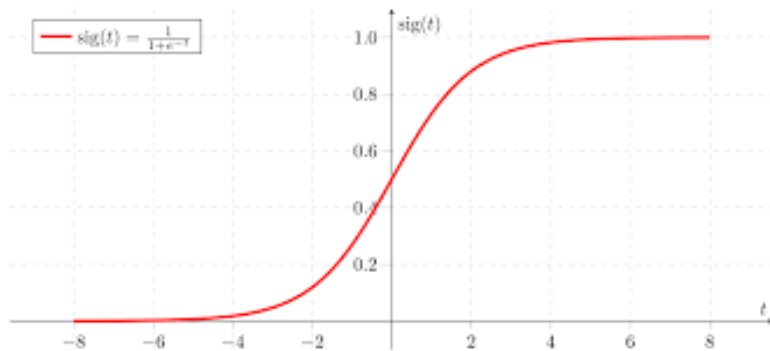
Optimización

$$\operatorname{argmin}_{a,b,c} -\frac{1}{N} \sum_{i=1}^N [Y_i \log(\hat{Y}(X_i, \theta)) + (1 - Y_i) \log(1 - \hat{Y}(X_i, \theta))]$$

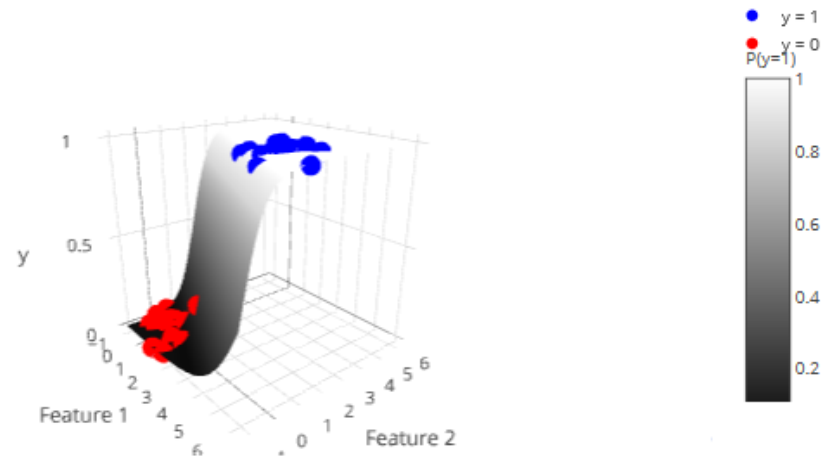


Aprendizaje supervisado → Ejemplo, clasificación

Regresión logística



Logistic Regression: 2 Features - Projected

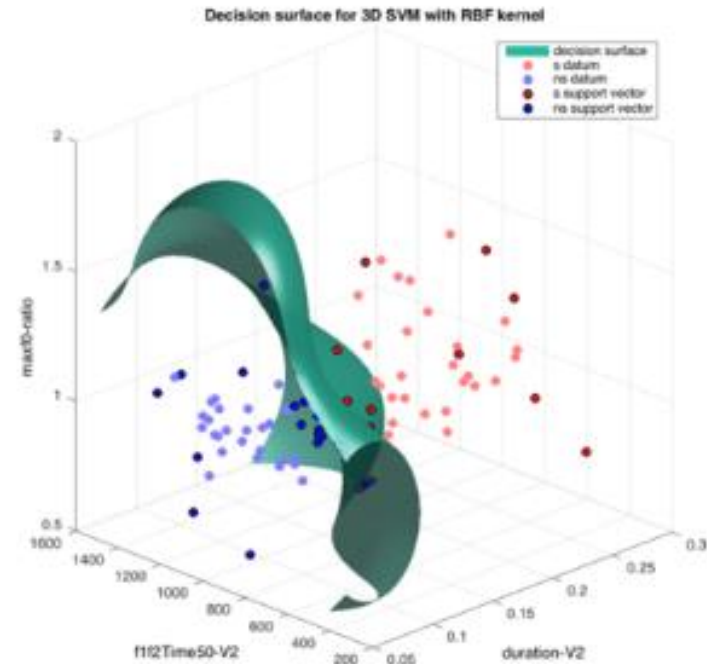
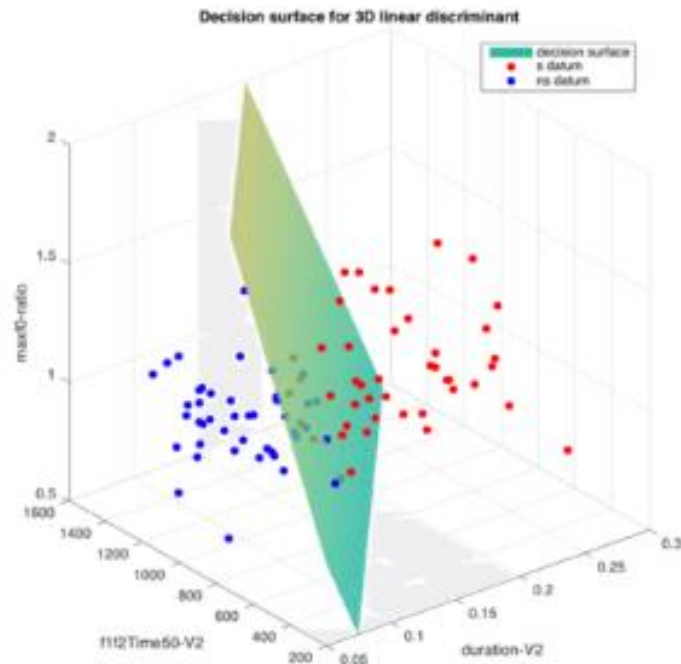


$$f(x) = \frac{1}{1 + e^{-(ax_1 + bx_2 + c)}}$$

Distancia del punto al plano

<https://florianhartl.com/logistic-regression-geometric-intuition.html>

<http://mathworld.wolfram.com/Point-PlaneDistance.html>



Superficies reales de discriminación



- Regresión logística es un método lineal (separación por planos)
- Es sencillo, pero en general funciona sorprendentemente bien
- Importancia de separar los datos en entrenamiento y test
- Importancia de la métrica de predicción (accuracy, recall, precisión,..)

Centro de computación I:

- Drive
- Tus ficheros en red



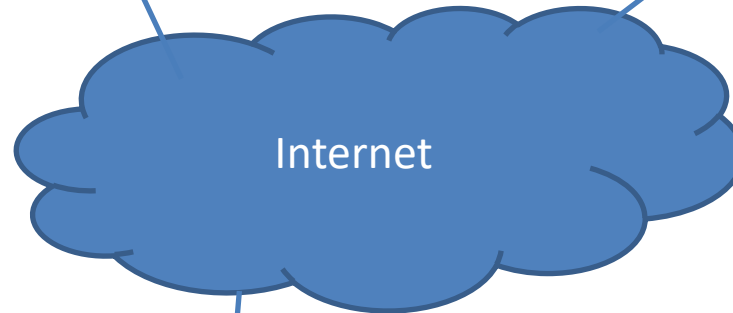
Código para conectar los dos entornos:

```
from google.colab import drive  
drive.mount('/content/drive')
```



Centro de computación II:

- Colaboratory
- El lugar donde se ejecutan los algoritmos ML



Tu ordenador

- Jupyter notebook