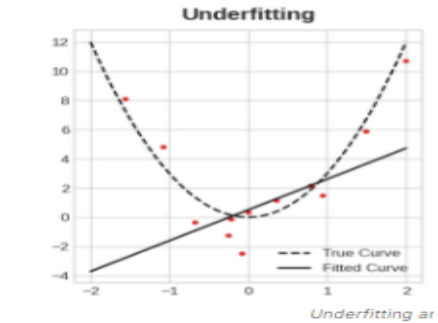
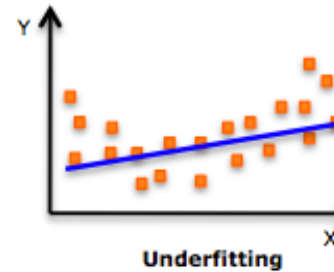
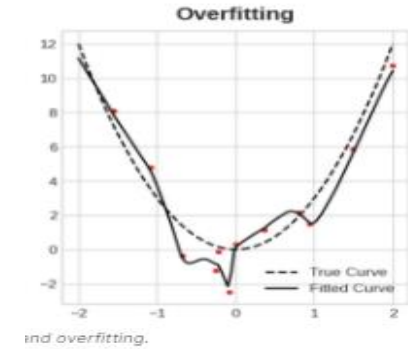
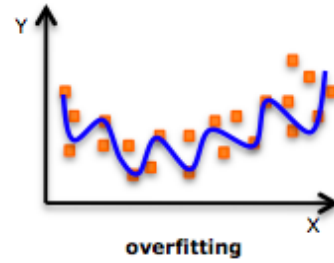


Introducción a la Ciencia de Datos



Lesson 3

Overfitting and underfitting

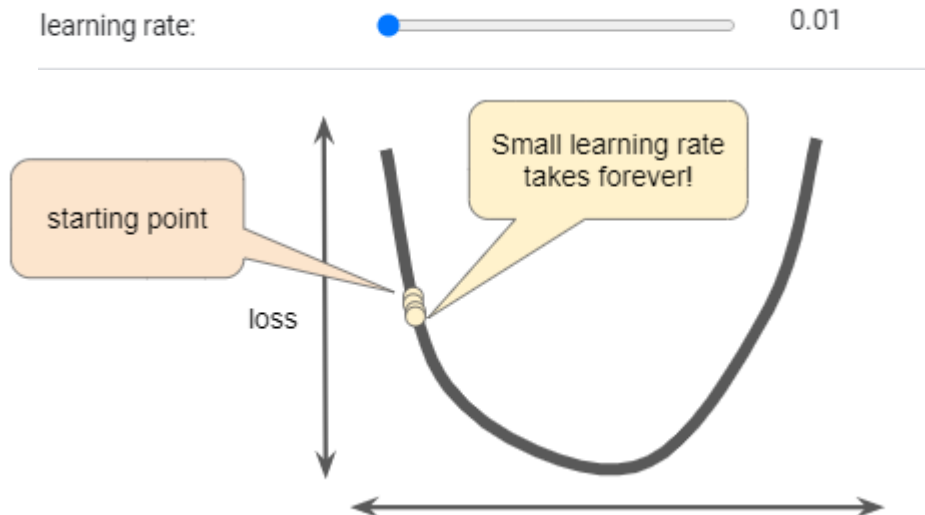


Lesson 4.1: Learning Rate

Epochs: Es una ronda completa de datos de entrenamiento

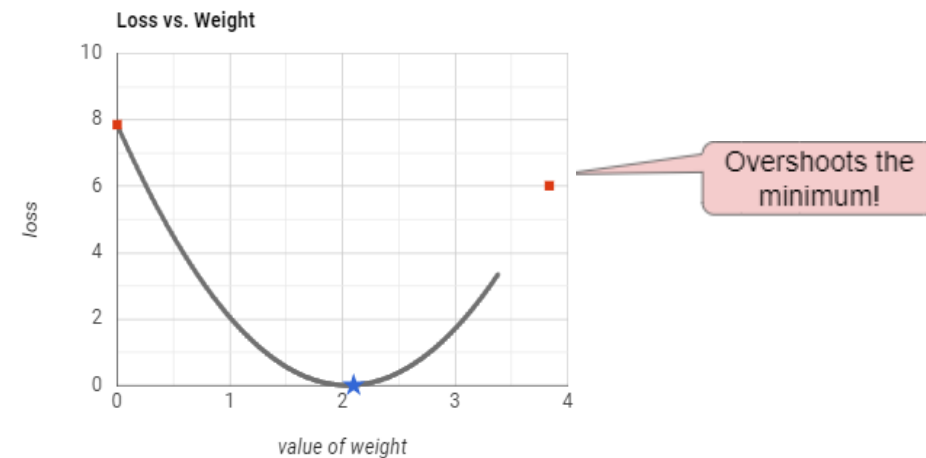
Learning Rate: La tasa de aprendizaje es un parámetro de ajuste que determina el tamaño del paso en cada iteración.

Por ejemplo: si la magnitud del gradiente es 2.5, y Learning Rate es 0.01, entonces determinará que el siguiente punto es 0.025 del punto anterior.



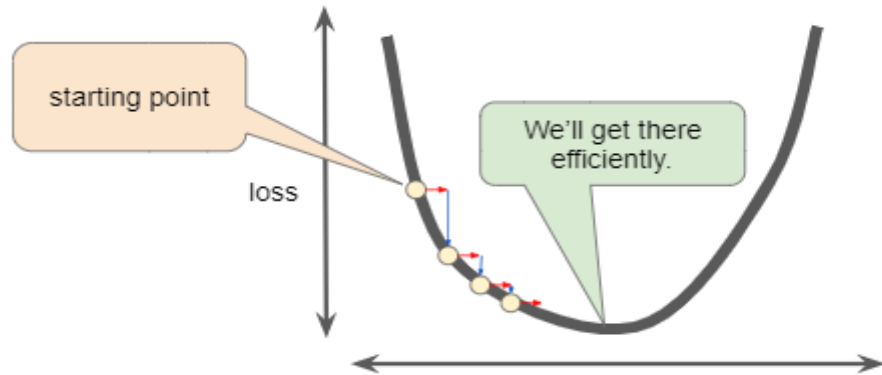
Los algoritmos de descenso de gradiente multiplican el gradiente por un escalar conocido como **Learning Rate** (también llamado **step size**) para determinar el siguiente punto.

Por ejemplo: si Learning Rate es igual a 0.5



Lesson 4.2: Interpreting the Learning Curves

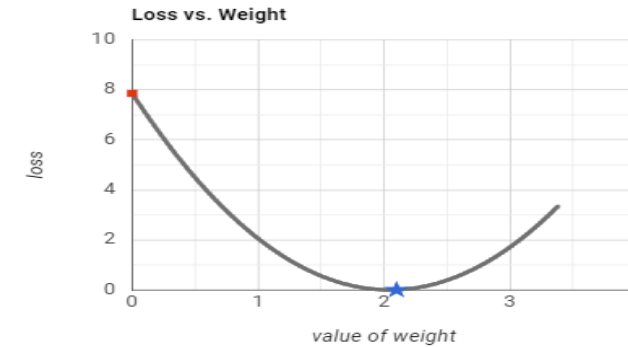
Figura 8. La tasa de aprendizaje es la correcta.



Set learning rate: 0.01

Execute single step: 0

Reset the graph:



Ejercicio 1: Establecer Learning Rate de 0.03. Siga presionando el botón STEP hasta que el algoritmo de descenso de gradiente alcance el punto mínimo de la curva de pérdida.

¿Cuántos pasos tomó?

Ejercicio 2: ¿Puede alcanzar el mínimo más rápidamente con una tasa de aprendizaje más alta? Establecer Learning Rate de 0.1 y siga presionando STEP hasta que el descenso del gradiente alcance el mínimo.

¿Cuántos pasos tomó esta vez?

Ejercicio 3: ¿Qué tal una tasa de aprendizaje aún mayor? Restablezca el gráfico, establecer Learning Rate de 1 e intente alcanzar el mínimo de la curva de pérdida.

¿Qué pasó esta vez?

Lesson 4.2: Overfitting

Se denomina sobreajuste (**overfitting**) al hecho de hacer un modelo tan ajustado a los datos de entrenamiento que haga que no generalice bien a los datos de test.

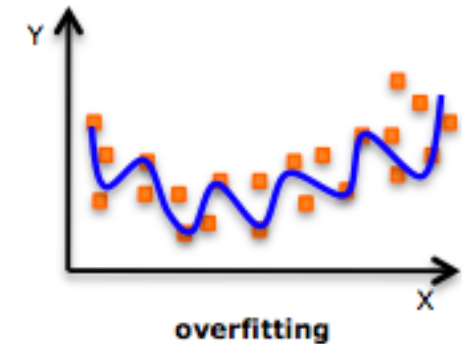
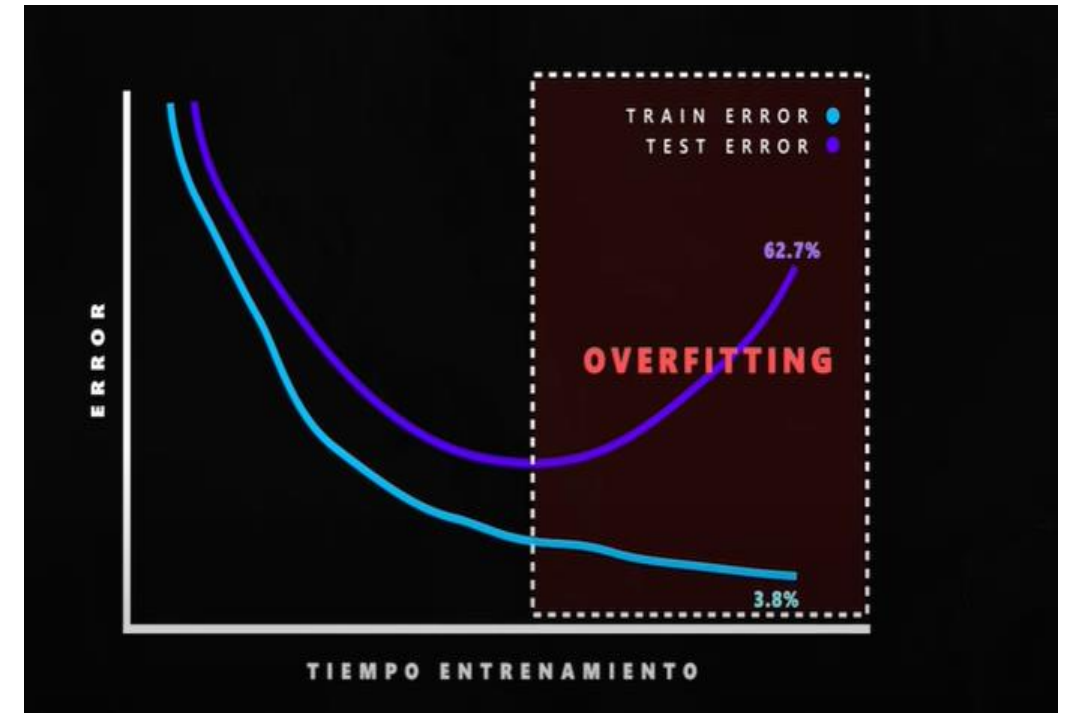
¿Cómo saber si hay overfitting?

Si el modelo entrenado con el conjunto de train tiene un 90% de aciertos y con el conjunto de test tiene un porcentaje muy bajo, esto señala claramente un problema de overfitting.

```
#dividir el dataset
from sklearn.model_selection import train_test_split

train_X, val_X, train_y, val_y = train_test_split(X, y, test_size=0.2, random_state=3)
```

Note: El truco para entrenar modelos de aprendizaje profundo es encontrar el mejor equilibrio entre los dos.



Lesson 4.3: Early Stopping

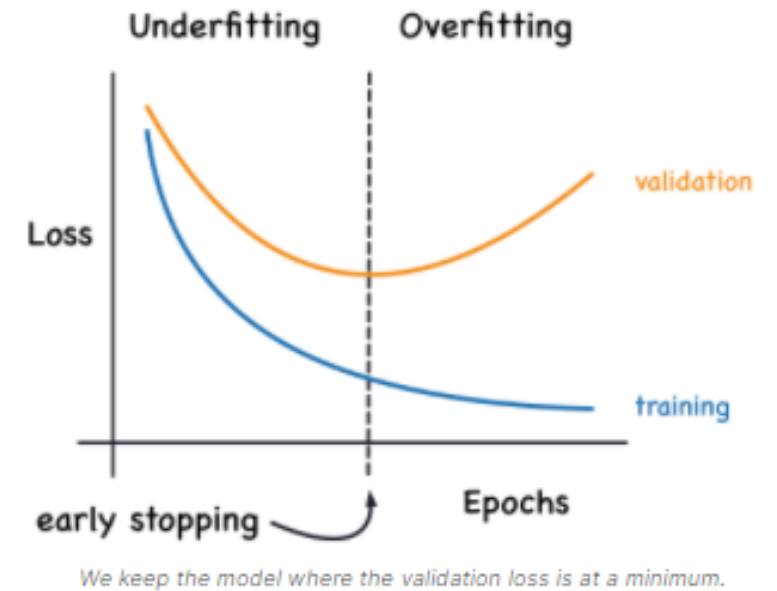
Early Stopping es una técnica de optimización que se utiliza para reducir el sobreajuste (**overfitting**) sin comprometer la precisión del modelo.

La idea principal de Early Stopping es detener el entrenamiento antes de que un modelo comience a sobreajustarse.

```
from tensorflow.keras import layers, callbacks
```

```
[ ] #early stopping para evitar el overfitting
```

```
early_stopping = callbacks.EarlyStopping(  
    min_delta=0.01, #cantidad mínima de cambio para contar como una mejora  
    patience=30,    #cuantas epocas esperar para detener el entrenamiento  
    restore_best_weights=True,  
)
```



Señal: Es la parte que se generaliza, la parte que puede ayudar a nuestro modelo a realizar predicciones a partir de nuevos datos.

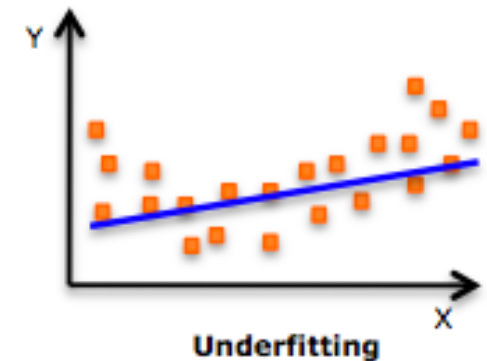
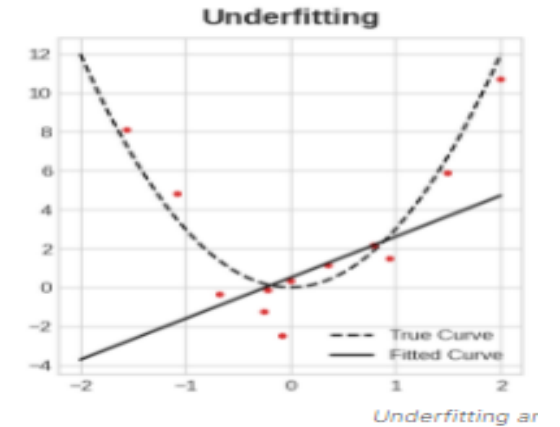
Ruido: El ruido es esa parte que solo es cierta en los datos de entrenamiento, el ruido es toda la fluctuación aleatoria que proviene de los datos que no pueden ayudar al modelo a hacer predicciones.

Lesson 4.6: Underfitting

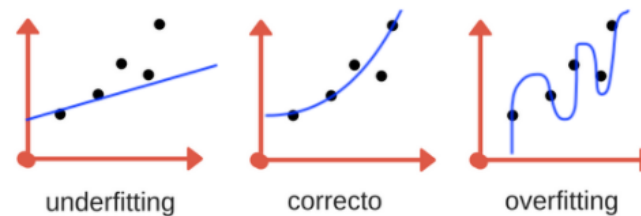
Underfitting (desajuste) es un escenario en Data Science en el que un modelo de datos no puede capturar la relación entre las variables del input y output con precisión, lo que genera una alta tasa de error tanto en el conjunto de entrenamiento como en los datos no vistos.

Formas de prevenir el **Underfitting**:

- Tratar los datos correctamente, eliminando los outliers y variables innecesarias.
- Aumentar las iteraciones en algunos casos (incrementar la duración del entrenamiento)
- Disminuir la **regularización**: Hay algunas técnicas que ayudan a reducir el ruido y los valores atípicos dentro de un modelo.



Lesson 4.7: Underfitting & Overfitting



Entreno al modelo con
1 sola raza de perro



Muestra nueva:
¿Es perro?



Entreno al modelo con
10 razas de perro color marrón

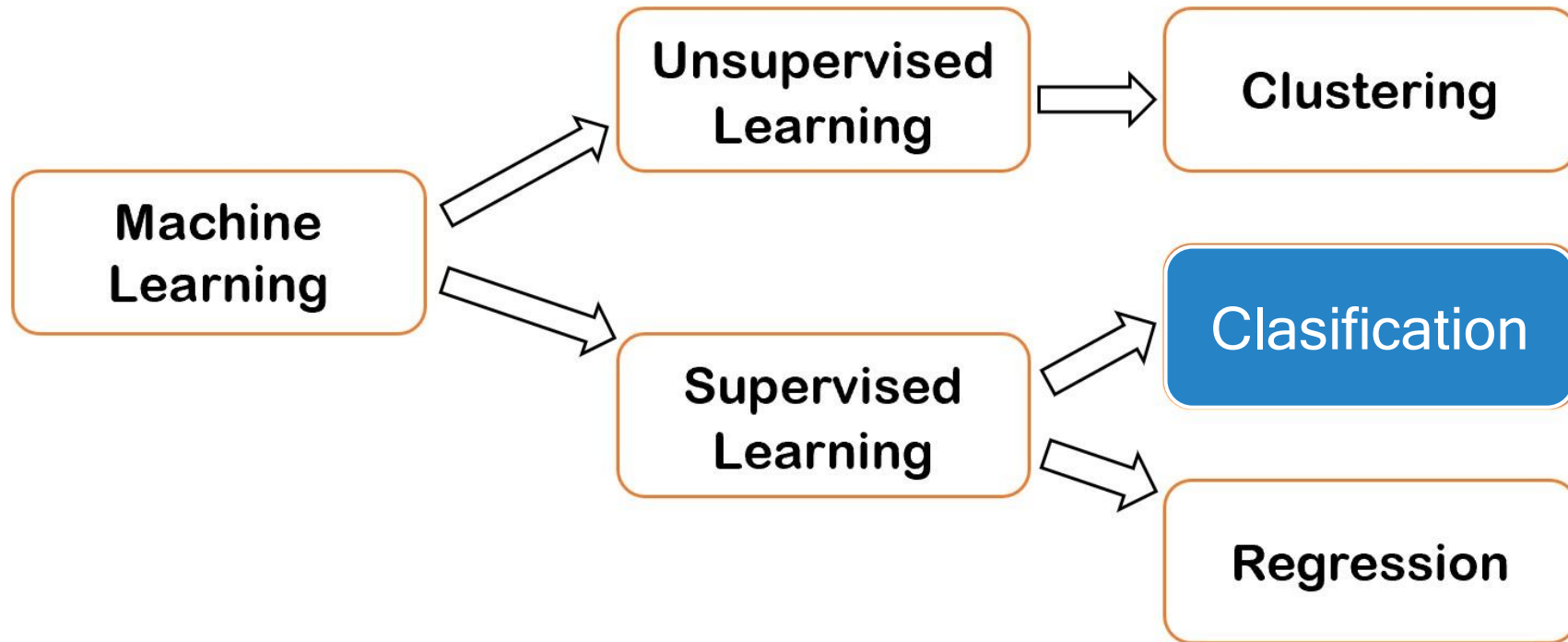


Muestra nueva:
¿Es perro?

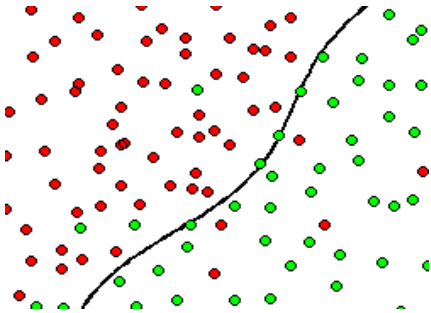


Aprendizaje Supervisado

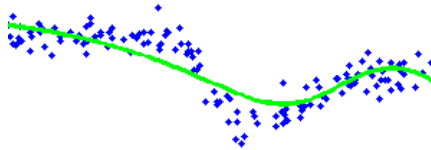
Modelos de Clasificación



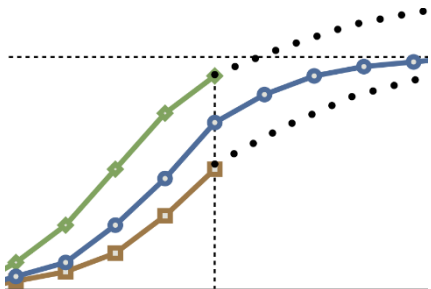
Tipos de Modelos Predictivos



- **Modelo de Clasificación** $p(Y=e) = f(X)$
Calcula la probabilidad de un evento
Asigna una instancia a una clase.

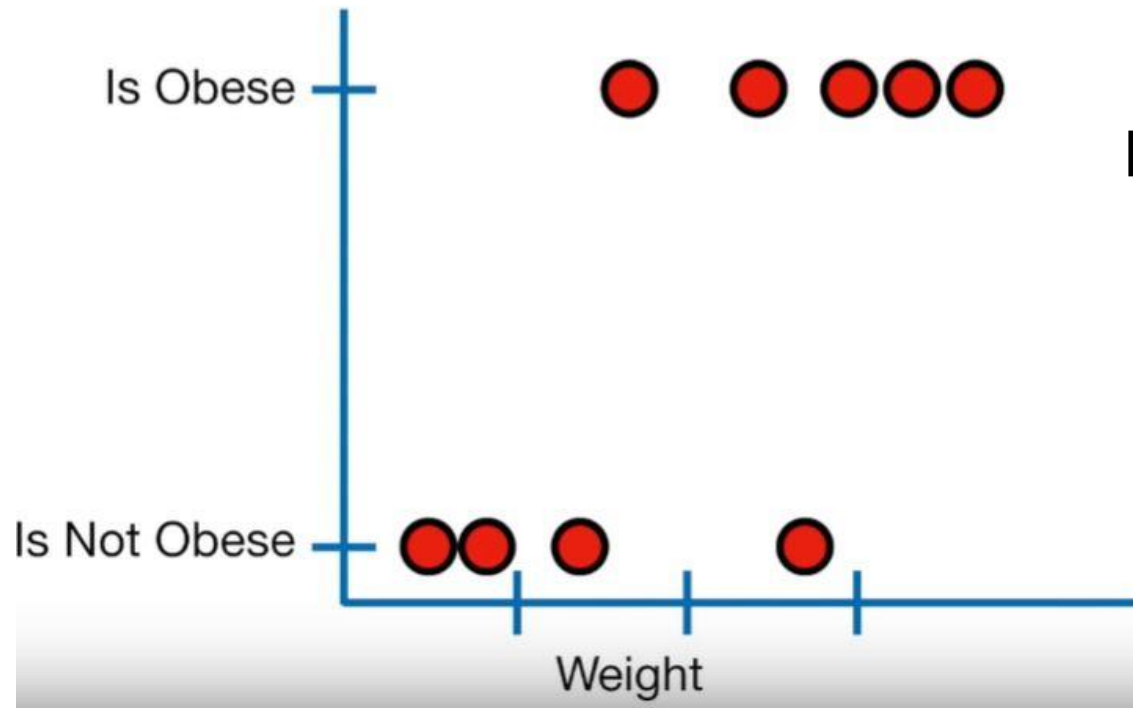


- **Modelo de Regresión** $Y = f(X)$
Calcula el valor puntual de una función
Es un problema de ajuste de curvas.



- **Modelo de Pronóstico** $Y_t = f(Y_{t-1}, X_{t-1}, \dots)$
Calcula el valor de una función en
diferentes momentos

¿Qué es lo que pasa si tenemos que predecir data **categorica**?



En caso tengamos:

- Una variable predictora (Weight)
- Debemos predecir si es obeso o no es obeso (**binario**)

Para estos casos utilizamos modelos de **clasificación**

Regresión Logística

Clasificación:

- En un problema de clasificación, los datos son clasificados en uno o más clases.
- Un problema de clasificación con dos clases es llamado binario, más de dos clases es llamado multiclase

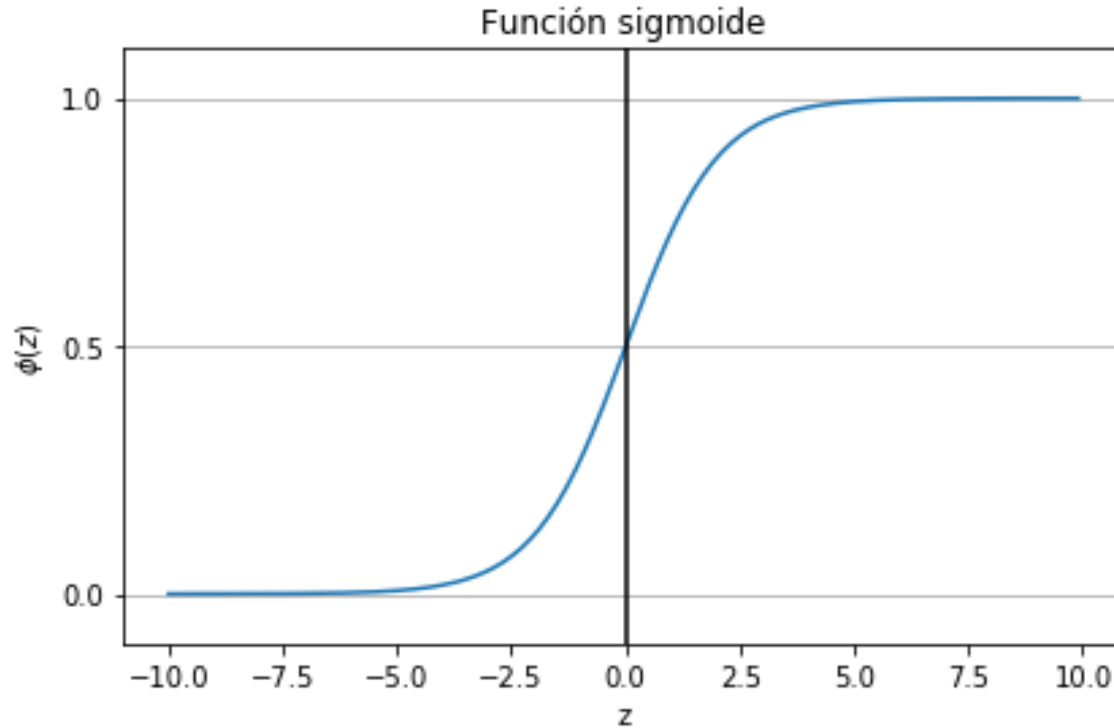


moroso



No moroso

Regresión Logística



El indicador que clasifica se denomina **punto de corte**:

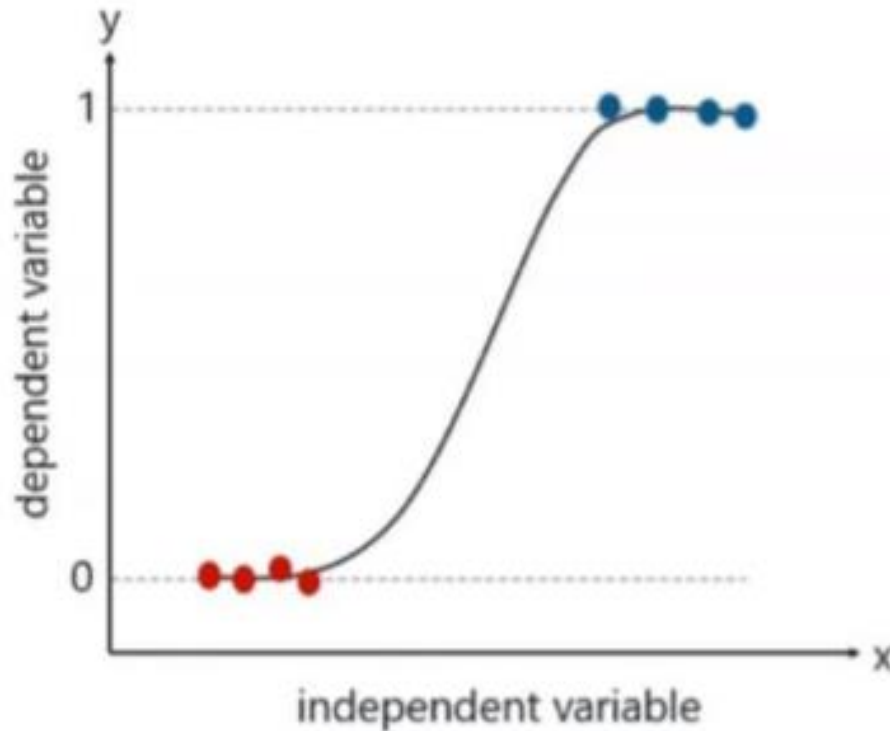
- Punto de corte (treehold) por defecto siempre es 0.5

$$Y_i = \begin{cases} 1 : \text{Evento exitoso} \\ 0 : \text{en caso contrario} \end{cases}$$

Para cada observación se calculará una **probabilidad** de obtener **cada uno de los valores del target**. Como el **target es binario** ambas probabilidades **suman 1**.

Si la probabilidad obtenida es **menor al punto de corte** entonces el modelo asignará el **target contrario**.

Regresión Logística



$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 x$$

- p es la probabilidad de que un evento ocurra, el cual estas tratando de predecir.
- X son variables independientes, las cuales determinan la ocurrencia del evento Y .
- b_0 es la constante, el cuál es la probabilidad de que el evento ocurra cuando los otros factores no son considerados