



Aprendizaje Supervisado - Clasificación

Clase sincrónica

Implementar modelos de aprendizaje automático por medio de técnicas estadísticas, adecuando los diferentes algoritmos debidamente a la situación y requerimientos necesarios

- Unidad 1: Introducción al Machine Learning
- Unidad 2: Aprendizaje Supervisado y No Supervisado
(Parte I: No supervisado)
(Parte II: Clasificación)
(Parte III: Clasificación)
(Parte IV: Regresión)
(Parte V: Series de tiempo)
- Unidad 3: Aplicando lo aprendido
(Parte I: Preprocesamiento de datos)
(Parte II: Modelamiento)



Te encuentras aquí



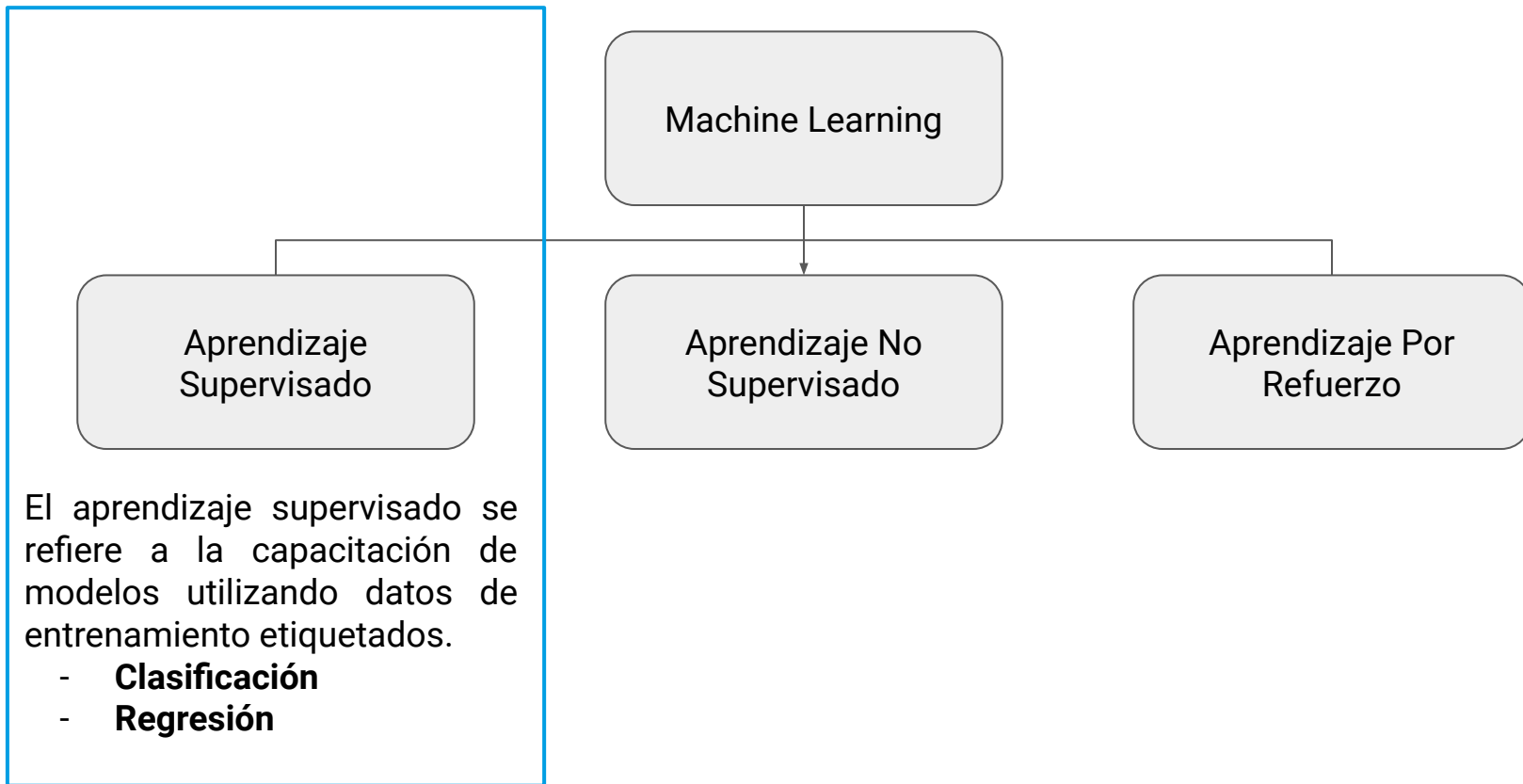
¿Qué aprenderás en esta sesión?

En esta sesión aprenderás sobre los algoritmos de aprendizaje supervisado, específicamente sobre algoritmos de clasificación. Al finalizar sabrás cuáles son, para qué sirven y cómo implementarlos.

¿Qué es el aprendizaje supervisado?



Aprendizaje Supervisado



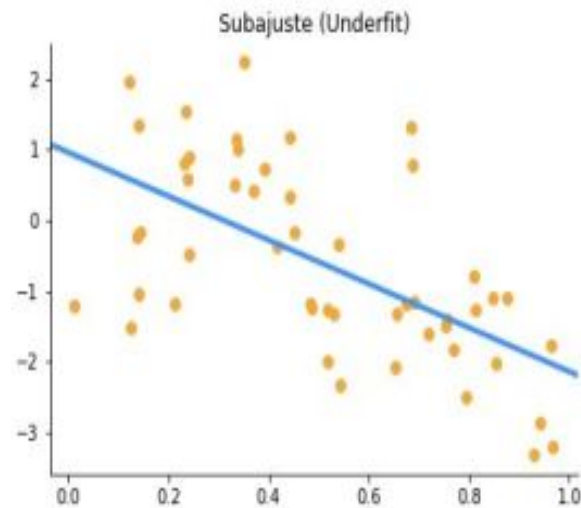
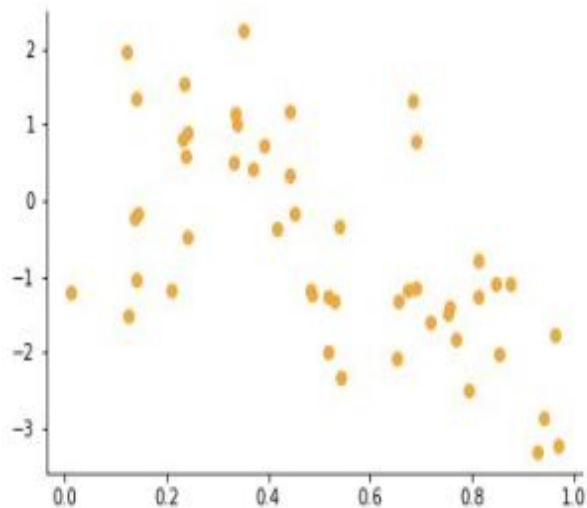
- ¿Qué algoritmos de clasificación recuerdan?
 - ¿Qué son los hiperparametros?
- ¿Cómo evaluamos un modelo de clasificación?



/* Overfitting y Underfitting */

Sobreajuste y Subajuste

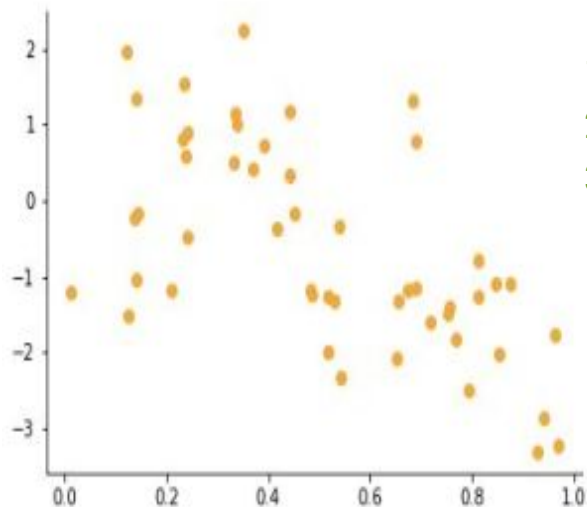
¿En qué consiste?



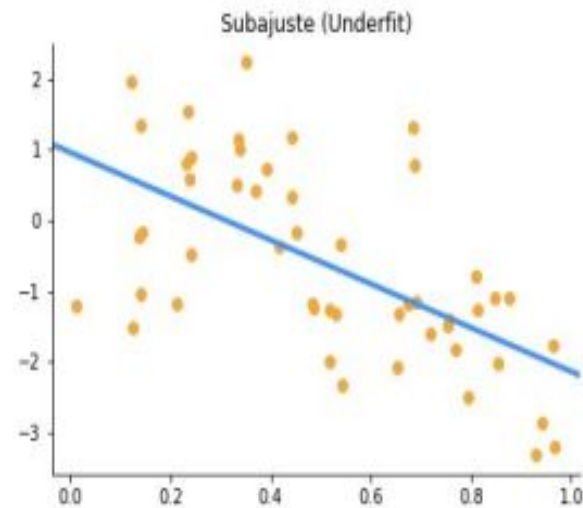
{desafío}
latam_

Sobreajuste y Subajuste

¿Qué problemas presenta?

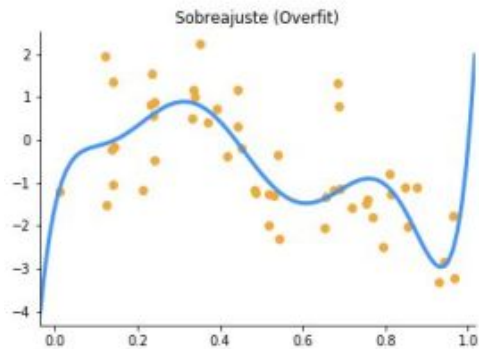
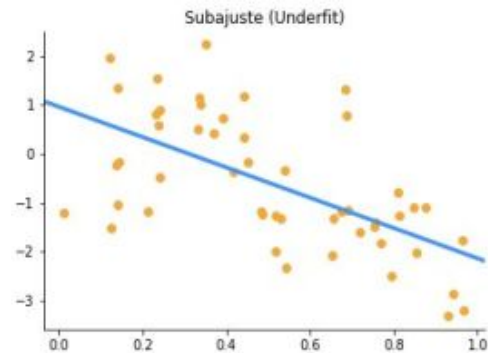
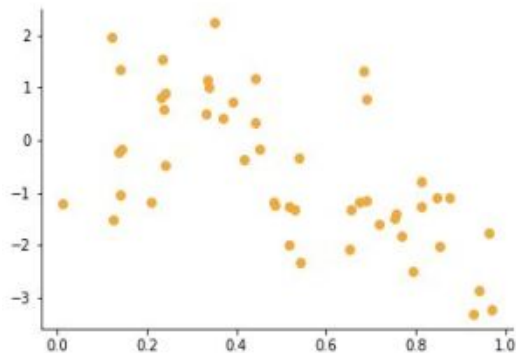


1. Alto Sesgo (forma inflexible)
2. Menor capacidad explicativa
3. Error generalizable



Sobreajuste y Subajuste

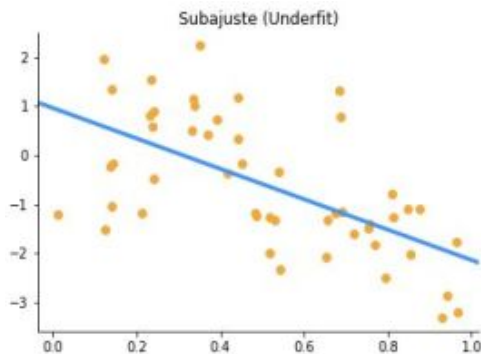
Expectativa v/s realidad



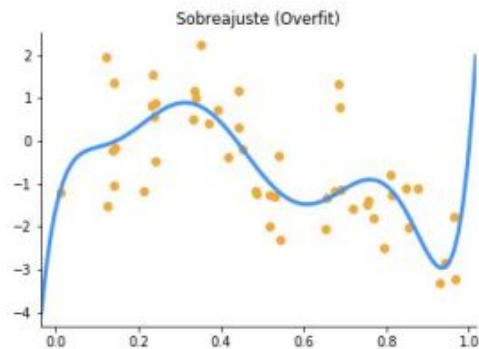
Sobreajuste y Subajuste

Comparación

1. Alto Sesgo (forma inflexible)
2. Menor capacidad explicativa
3. Error generalizable



1. Alto Varianza (forma acoplada)
2. Mayor capacidad explicativa
3. Error poco generalizable



/* Regresión Logística */

Regresión Logística

¿En qué consiste?

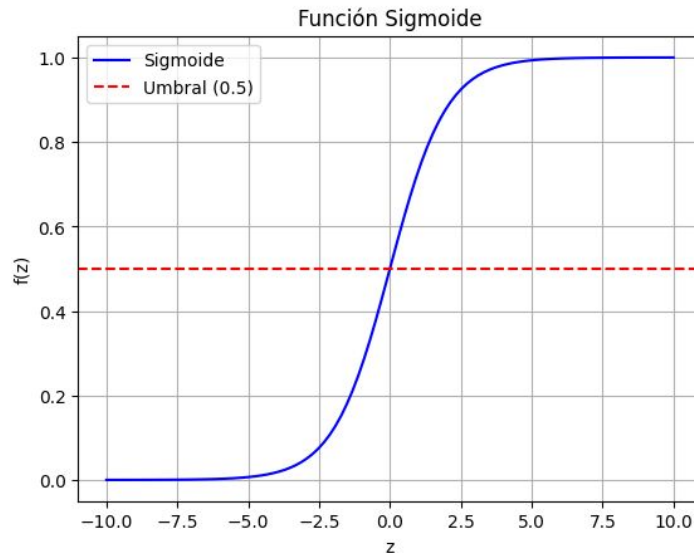
Modelo de clasificación que utiliza la **función logística** (función sigmoide) para predecir la probabilidad de que una observación pertenezca a una clase.

$$f(z) = \frac{1}{1+e^{-z}}$$

$$z = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

Probabilidad de pertenecer a la clase 1

$$P(y = 1|X) = f(z)$$



Regresión Logística

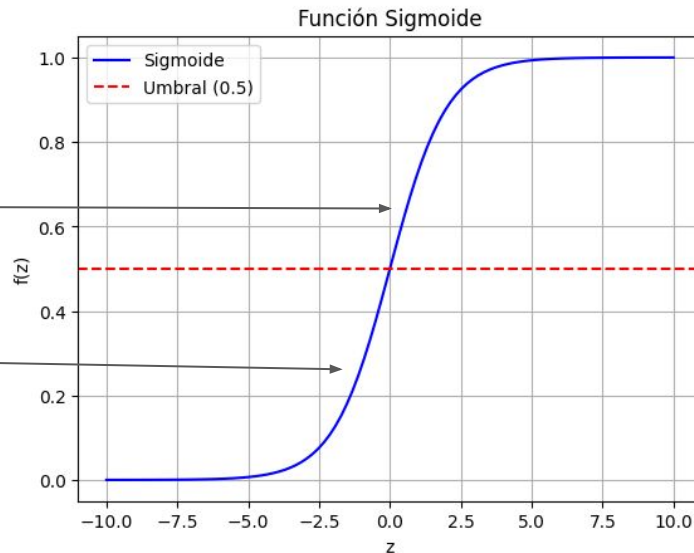
¿En qué consiste?

Modelo de clasificación que utiliza la **función logística** (función sigmoide) para predecir la probabilidad de que una observación pertenezca a una clase.

Probabilidad de pertenecer a la clase 1

$$P(y = 1|X) = f(z)$$

$$P(0) = 1 - P(y = 1|X)$$



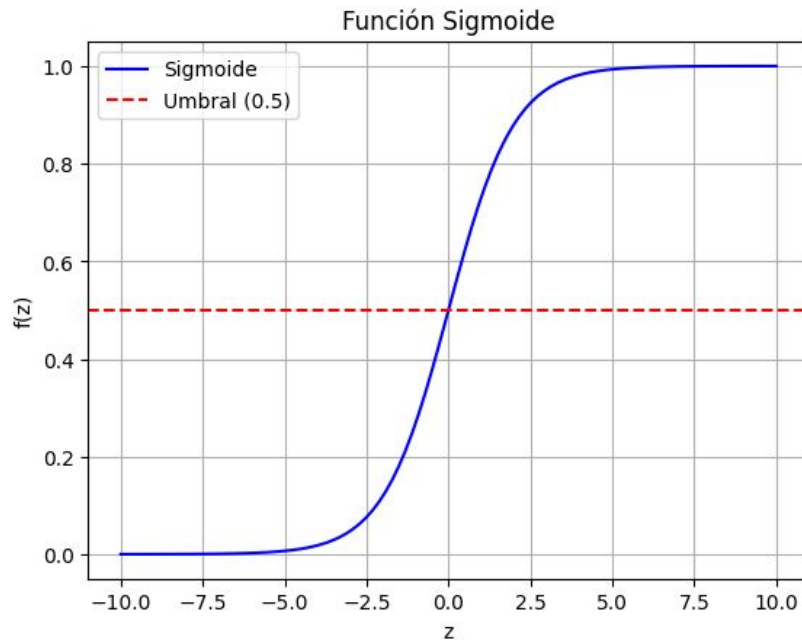
Regresión Logística

Determinando los parámetros

$$P(y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\dots+\beta_nx_n)}}$$

¿Cómo encontramos los β que ajusten mejor la función a la realidad?

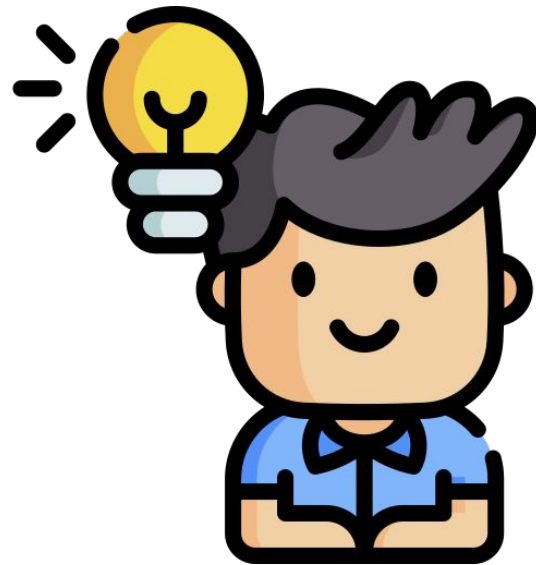
¡Método de Máxima Verosimilitud!



Ecuación de Máxima Verosimilitud

Idea básica

Encontrar los parámetros que maximizan la ocurrencia de los datos de entrenamiento.



Ecuación de Máxima Verosimilitud

¿Cómo se aplica?

Función de probabilidad conjunta

La probabilidad conjunta de los datos observados es la multiplicación de las probabilidades individuales de los datos.

$$P(Y_i = 1|X) = p_i$$

$$L(\beta) = \prod P(Y_i = 1|X)$$

$$\log(L(\beta)) = \sum \log(P(Y_i|X_i; \beta))$$

¡Veamos con Python!



Regresión Logística con Python

Aplicando nuestro código

Vamos a ver de manera práctica la implementación de la regresión logística con Python. Para ello, puedes abrir tu propio archivo y seguir los pasos que te mostrará tu profesor en la presentación. En esta abordaremos:

1. Regresión logística
2. Support vector machine
3. Hiperparámetros
4. Métricas - ROC AUC



Desafío

“Clasificación de calidad de ramen”

- Descarga el archivo “Desafío”.
- Tiempo de desarrollo asincrónico: desde 4 horas.
- Tipo de desafío: individual.

¡AHORA TE TOCA A TI! 💪



Ideas fuerza



Los conceptos de **Overfitting** y **Underfitting** son muy importantes para reconocer la generalización de los modelos.

Aprender algoritmos muy importantes para la clasificación como son **regresión logística** y **support vector machine**.

El **área bajo la curva ROC (AUC ROC)** es una métrica muy relevante para los modelos de clasificación que no es susceptible al desbalanceo de clases.

¿Qué conceptos no te
quedaron claros o quieres
reforzar?



Recursos asincrónicos

¡No olvides revisarlos!

Para esta semana deberás revisar:

- Guía de estudio
- Desafío “Clasificando los diferentes ramen”





Próxima sesión...

*Veremos modelos de aprendizaje supervisado,
específicamente estudiaremos modelos de regresión en
conjunto con las métricas correspondientes.*

{desafío}
latam_

*Academia de
talentos digitales*

