

1. Introducción al machine learning.

En el contexto del aprendizaje supervisado, se trabaja con variables independientes (también conocidas como características, predictores o variables de entrada) y variables dependientes (también llamadas variables objetivo o target o variables de salida). Las variables independientes son las que se utilizan para hacer predicciones o estimaciones, mientras que las variables dependientes son los resultados que se intentan predecir o modelar. Por ejemplo, en un problema de predicción de precios de viviendas, las características como el número de habitaciones, la ubicación y el tamaño de la propiedad serían las variables independientes, mientras que el precio de venta sería la variable dependiente.

- Variables Independientes (entradas)

1. Cualitativas

- a) Texto

- Nominales (categorías, Por ejemplo: países, sexo)
- Ordinales (poco, mucho, muchísimo, Por ejemplo: nivel de tabaquismos)

- b) Numéricas

- Nominales
- Ordinales

2. Cuantitativas: cuando hablamos de cantidad

- a) Discretas: Por ejemplo año, mes, edad, etc.
- b) Continuas: Por ejemplo altura, peso, etc.

- Variables Dependientes (salidas, categorías)

Las variables **cualitativas** son aquellas que describen características o cualidades y no pueden ser medidas en términos numéricos. Las variables **cuantitativas**, por otro lado, son aquellas que pueden ser medidas en términos numéricos y tienen valores numéricos.

Las **variables nominales** son aquellas que no tienen un orden natural, como el género o el color de ojos. Las **variables ordinales** son aquellas que tienen un orden natural, como el nivel de educación (primaria/secundaria/universidad), nivel de tabaquismos: Clasificamos como leve (1), moderado (2), nivel medio (3), importante (4) y muy importante (5).

Variables y tipos de problemas (video 02a min 9:00)

1. Si la variable dependiente es **cualitativa**, el tipo de problema es de **clasificación**.
2. Si la variable dependiente es **cuantitativa**, el problema es de **regresión**. Por ejemplo si quiero predecir el precio de una propiedad.

3. Si **NO hay variable** dependientes, el problema es agrupamiento.
- **Outliers:** Valores atípicos, pueden ser errores o un dato que se sale de la norma.
 - **Correlación:**
 - **Positiva:** Cuando una variable aumenta la otra también.
 - **Negativa:** Cuando una variable aumenta la otra disminuye.
 - **Sin correlación:** Cuando una variable aumenta la otra no cambia.
 - **Varianza:** Es la medida de dispersión de una variable respecto a su media. Si la varianza es alta, los datos están muy dispersos, mientras que si la varianza es baja, los datos están muy agrupados.
 - **Covarianza:** Es una medida de la relación lineal entre dos variables aleatorias. Indica cómo varían conjuntamente dos variables aleatorias respecto a sus medias. Si la covarianza es positiva, las variables aumentan o disminuyen conjuntamente, mientras que si la covarianza es negativa, una variable aumenta mientras la otra disminuye.