Regresión Lineal

**Parte 1:**

Imagina que eres un analista de datos en una empresa de ventas de electrodomésticos. La gerencia te ha proporcionado un conjunto de datos que contiene información sobre las ventas mensuales de varios productos durante un período de varios años. Tu tarea es utilizar técnicas de regresión para analizar los datos y predecir las ventas futuras en función de ciertas variables independientes. A continuación, se presentan algunas preguntas para que las investigues y respondas utilizando los conceptos relacionados con la regresión.

**Preguntas:**

**Construcción del Modelo:**

* ¿Qué es la regresión lineal y cómo se utiliza para predecir valores?
* Explica el proceso de construcción de un modelo de regresión lineal.
* ¿Qué son las variables independientes y la variable dependiente en el contexto de la regresión?

**Evaluación del Modelo:**

* Describe las métricas comunes utilizadas para evaluar la precisión de un modelo de regresión.
* ¿Qué es el coeficiente de determinación (R²) y qué información proporciona sobre el modelo?

**Supuestos de la Regresión:**

* Menciona los supuestos básicos de la regresión lineal y por qué son importantes.
* ¿Qué es la multicolinealidad y cómo puede afectar los resultados de un modelo de regresión?

***Importante*: Se deberá investigar sobre los tópicos que se tratan en el caso de ser necesario.**

**Parte 2**

Imagina que eres un científico de datos que está trabajando en un proyecto para predecir el precio de las casas en una determinada área en función de varias variables como el tamaño de la casa, el número de habitaciones, la ubicación, etc.

**Escenario:**

Un agente inmobiliario quiere predecir el precio de las casas en una ciudad basándose en el tamaño de la casa en pies cuadrados. Ha recopilado datos de 10 casas y sus respectivos precios en miles de dólares. Los datos son los siguientes:

| **Tamaño de la Casa (pies cuadrados)** | **Precio (en miles de dólares)** |
| --- | --- |
| 1500 | 250 |
| 2000 | 300 |
| 2500 | 350 |
| 3000 | 400 |
| 3500 | 450 |
| 4000 | 500 |
| 4500 | 550 |
| 5000 | 600 |
| 5500 | 650 |
| 6000 | 700 |

**Preguntas:**

* ¿Cuál es la variable independiente en este escenario?
* ¿Cuál es la variable dependiente en este escenario?
* ¿Qué tipo de relación esperas encontrar entre estas dos variables?
* ¿Cuál es la ecuación de la recta de regresión que mejor se ajusta a estos datos (simple/multiple)?

**Parte 3**

**Ejercicio 1: Entendiendo la Regresión Lineal Simple**

Considera el siguiente conjunto de datos:

| **X** | **Y** |
| --- | --- |
| 1 | 3 |
| 2 | 5 |
| 3 | 7 |
| 4 | 9 |
| 5 | 11 |

Utilizando la regresión lineal simple, encuentra la ecuación de la recta de mejor ajuste para estos datos.

**Ejercicio 2: Coeficiente de Determinación (R²)**

Considera el siguiente conjunto de datos para una regresión lineal simple:

| **X** | **Y** |
| --- | --- |
| 1 | 2 |
| 2 | 4 |
| 3 | 5 |
| 4 | 4 |
| 5 | 6 |

Calcula el coeficiente de determinación (R²) para estos datos.

**Generalidades**

Para la resolución de la práctica, puede utilizar Excel. Posteriormente veremos ejemplos de aplicación en Azure ML.

**Links de Interés:**

* <https://www.youtube.com/watch?v=gZnT4_Wnb4s>
* <https://www.youtube.com/watch?v=lhK4Nl1Eiug>