

Regresión lineal simple.

▼ Desarrollo matemático para Regresión Lineal Simple

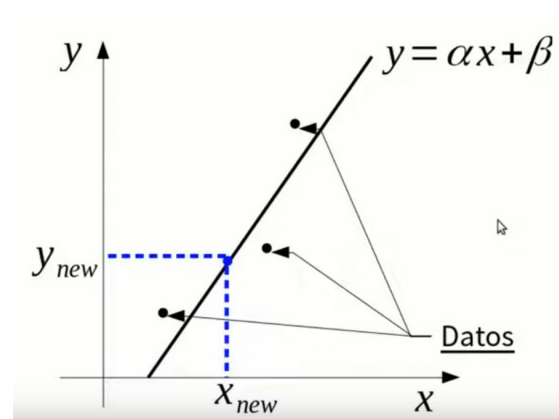
La regresión lineal es un modelo matemático que relaciona de forma lineal una variable dependiente con una o más variables independientes.

En el caso de una única variable independiente la ecuación es la siguiente:

$$y = \alpha \cdot x + \beta + \epsilon$$

Siendo:

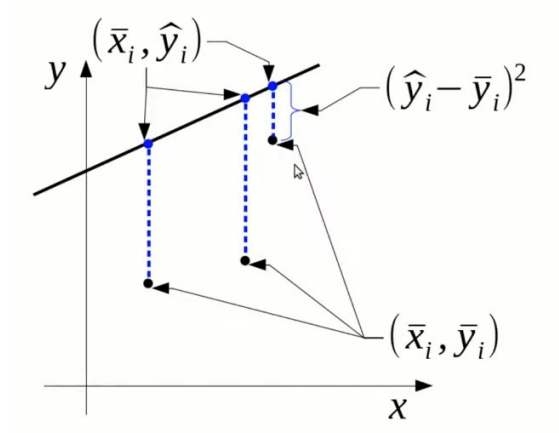
- α la pendiente de la recta
- β la ordenada en el origen
- ϵ el posible error cometido en la modelización



▼ Concepto de Error

Definimos el error como la diferencia entre el valor que predecimos al usar el modelo lineal que hemos creado y el valor real que obtenemos de los datos.

Podemos crear distintas funciones matemáticas para cuantificar el error.



▼ Funciones de error

Mean Absolute Error (MAE):

- Podemos medir directamente la diferencia en valor absoluto (distancia en el eje y) entre el valor real y el predicho y usar la media de esa medida como “Función de error”. Ese sería el Error absoluto medio (MAE Mean Absolute Error)

- **Fórmula:**

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- Donde n es el número de observaciones, y_i es el valor real, y \hat{y}_i es la predicción.

- **Interpretación:**

- Cuanto menor sea el MAE, mejor. Indica que las predicciones del modelo están más cerca de los valores reales. Los errores más grandes tienen un impacto lineal en el MAE.

Mean Squared Error (MSE):

- También podemos usar el valor elevado al cuadrado, para penalizar las distancias (errores) más altos y dar más peso en el cálculo del error a los valores más alejados de la recta. Ese sería el Error Cuadrático Medio (MSE Mean Square Error)

- **Fórmula:**

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- Donde n es el número de observaciones, y_i es el valor real, y \hat{y}_i es la predicción.

- **Interpretación:**

- Cuanto menor sea el MSE, mejor. Indica que las predicciones del modelo están más cerca de los valores reales. Los errores más grandes tienen un impacto cuadrático en el MSE.

Comparación:

- **Sensibilidad a Outliers:**

- El MSE penaliza más fuertemente los errores grandes debido al término cuadrático. Por lo tanto, es más sensible a outliers que el MAE.
- **Robustez a Outliers:**
 - El MAE es más robusto a outliers porque no eleva los errores al cuadrado. Los errores grandes tienen un impacto constante en el MAE.
- **Interpretación:**
 - Ambos indicadores miden la calidad de las predicciones, pero la elección entre MSE y MAE depende del énfasis en la sensibilidad a errores grandes y la robustez a outliers en el problema específico.

R al cuadrado R^2

El coeficiente de determinación R^2 mide la proporción de la variabilidad en la variable dependiente que es explicada por el modelo. Su valor varía entre 0 y 1, y en algunos casos, puede ser negativo.

- **Mejor Posible Valor:**
 - El mejor valor posible de R^2 es 1.0. Esto indicaría que el modelo explica el 100% de la variabilidad en la variable dependiente, es decir, todas las observaciones caen exactamente en la línea de regresión.
- **Valor Nulo:**
 - Un R^2 de 0.0 indica que el modelo no explica ninguna variabilidad en la variable dependiente. Podría interpretarse como si el modelo es tan malo como un modelo constante que siempre predice el valor medio de la variable dependiente, sin importar los valores de las variables independientes.
- **Valor Negativo:**
 - En algunos casos, especialmente cuando se utilizan modelos muy malos, el R^2 puede ser negativo. Esto significa que el modelo es tan pobre que incluso un modelo constante sería mejor.

Entonces, si el modelo siempre predice el valor de y , el mejor valor posible de R^2 debería ser 1.0, ya que eso indicaría que el modelo explica el 100% de la variabilidad en la variable dependiente. Si el R^2 es 0.0, significa que el modelo no está proporcionando ninguna mejora sobre un modelo constante que siempre predice el valor medio de y . La interpretación es relativa al rendimiento del modelo en comparación con un modelo nulo o constante.