Aprendizaje Supervisado: Regresión

Formación de Talento Altamente Especializado

Aprendizaje Supervisado: Regresión

Definición

Regresión: mecanismo estadístico para encontrar una relación numérica contínua basado en una o más variables, por medio de una ecuación o un conjunto de reglas (algoritmo).

Algoritmos de Regresión

Redes neuronales

Algoritmos basados en árboles

Máquinas de soporte vectorial

Regresión Lineal

Naive Bayes

K-Vecinos más cercanos

Algoritmos de Regresión

- Redes neuronales
- Algoritmos basados en árboles
- Máquinas de soporte vectorial
- Regresión Lineal
- Naive Bayes
- K-Vecinos más cercanos

Algoritmos de Regresión

Regresión Lineal

- Simplicidad
- Velocidad
- Entendimiento

Regresión Lineal

Definición

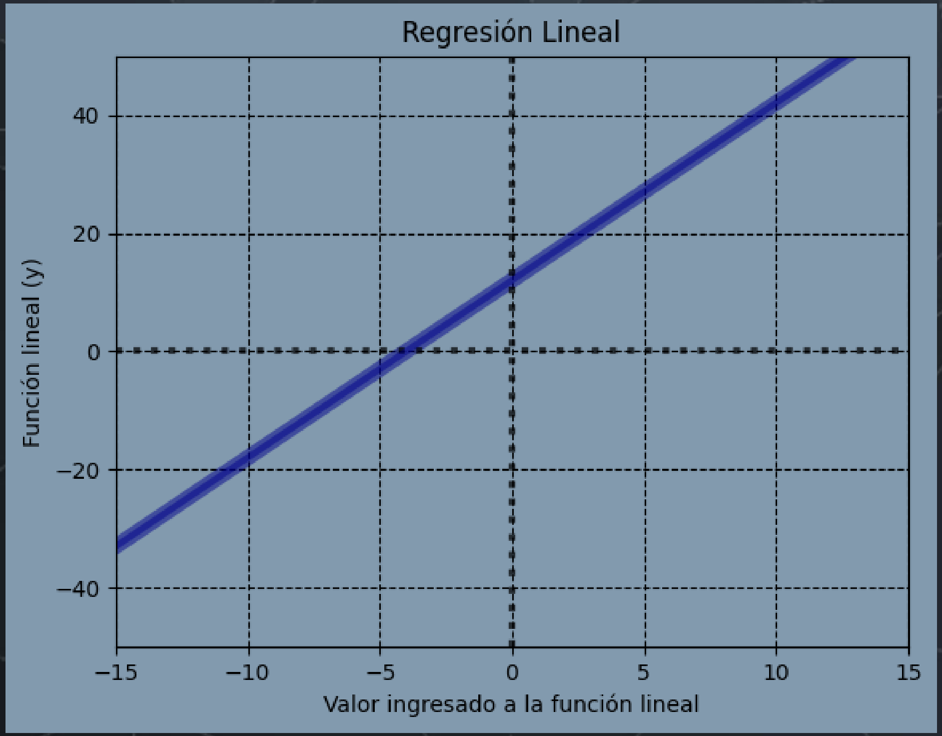
Regresión Lineal: Técnica utilizada para predecir un valor continuo, basado en una o más variables independientes.

Detalles

Asume una relación lineal entre las variables independientes y la dependiente.

Regresión Lineal

$$y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_nX_n$$



Regresión Lineal

$$y = \underbrace{\underbrace{a_0 + a_1X_1 + a_2X_2 + \ldots + a_nX_n}_{\text{}}}_{\text{}}$$

La solución consiste en encontrar el coeficiente a_i, asociado a la variable X_i. Lo encontraremos este para cada feature que tengamos disponible.

Se calcula las multiplicaciones y sumas. Así obtenemos el valor de y el cual es un valor continuo

Regresión Lineal

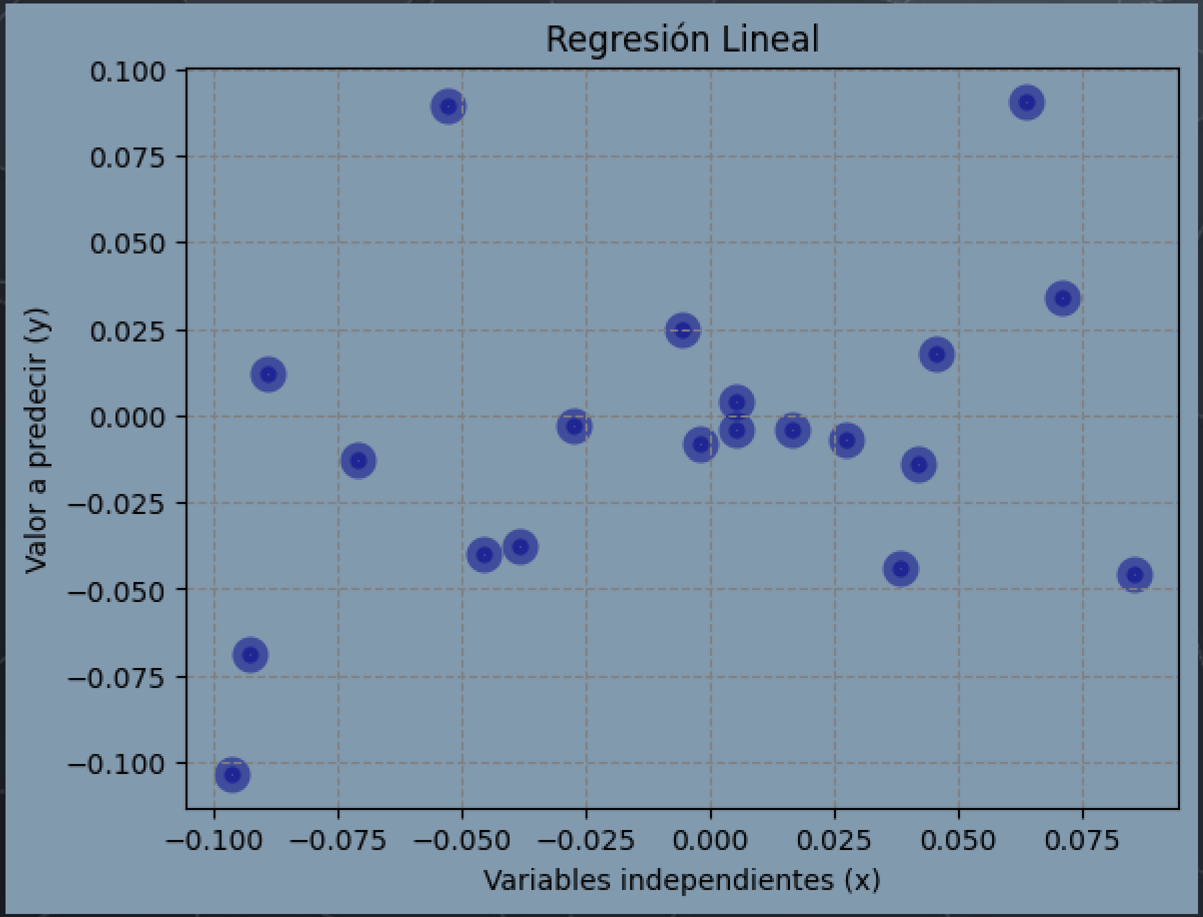
¿Pero cómo? Mínimos cuadrados.

Procedimiento matemático de optimización donde se busca minimizar el error (al cuadrado) de los datos a una función lineal.

Se requiere: Cálculo y Algebra Lineal

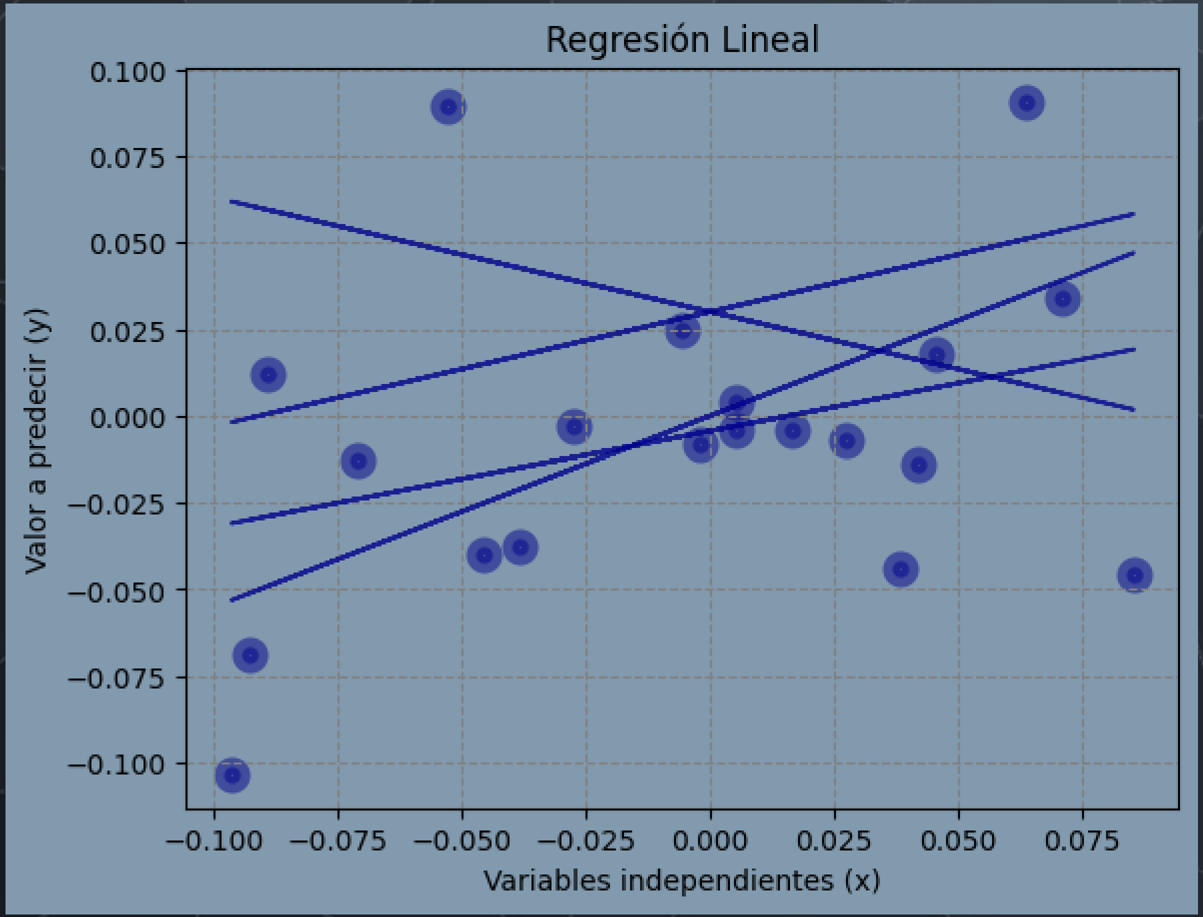
Regresión Lineal

¿Pero cómo? Mínimos cuadrados.



Regresión Lineal

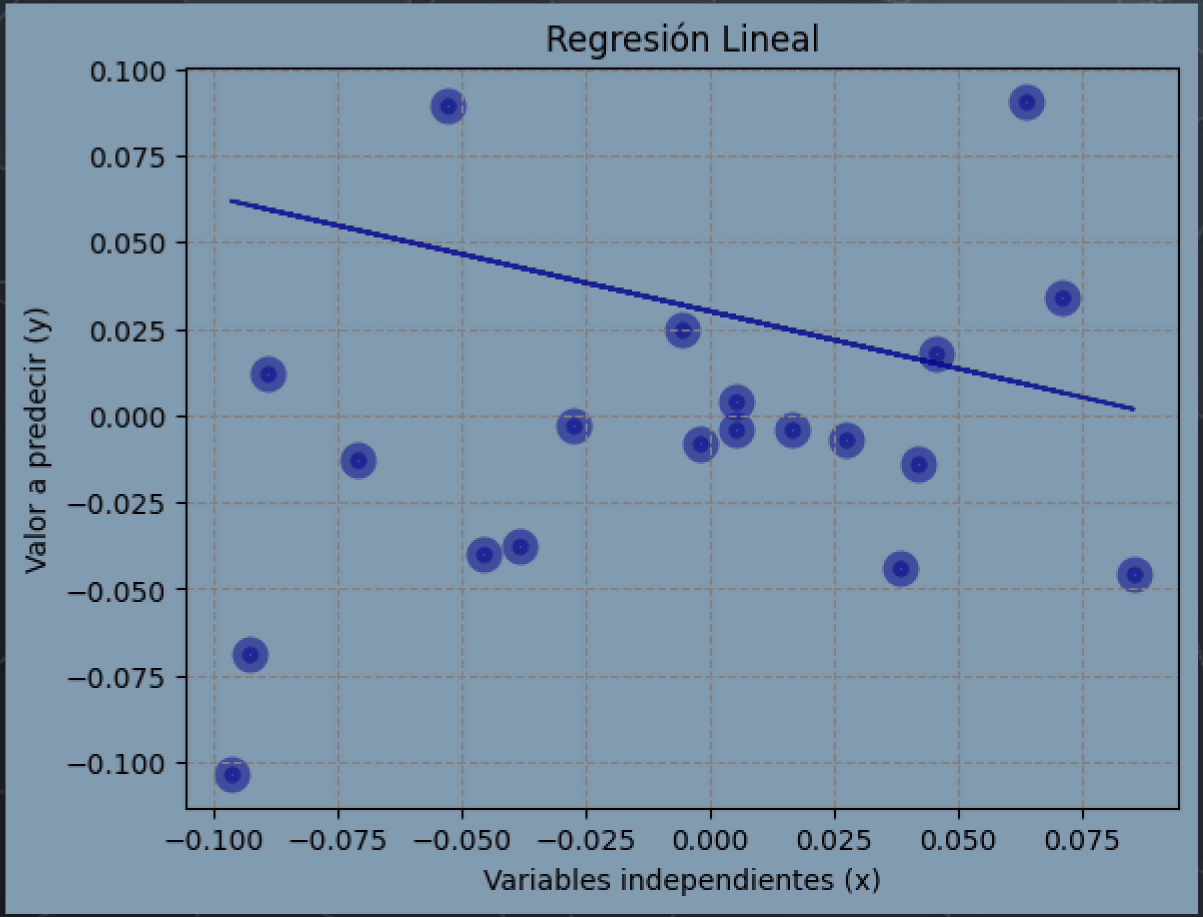
¿Pero cómo? Mínimos cuadrados.



TAE IA
Formación de Talento Altamente Especializado

Regresión Lineal

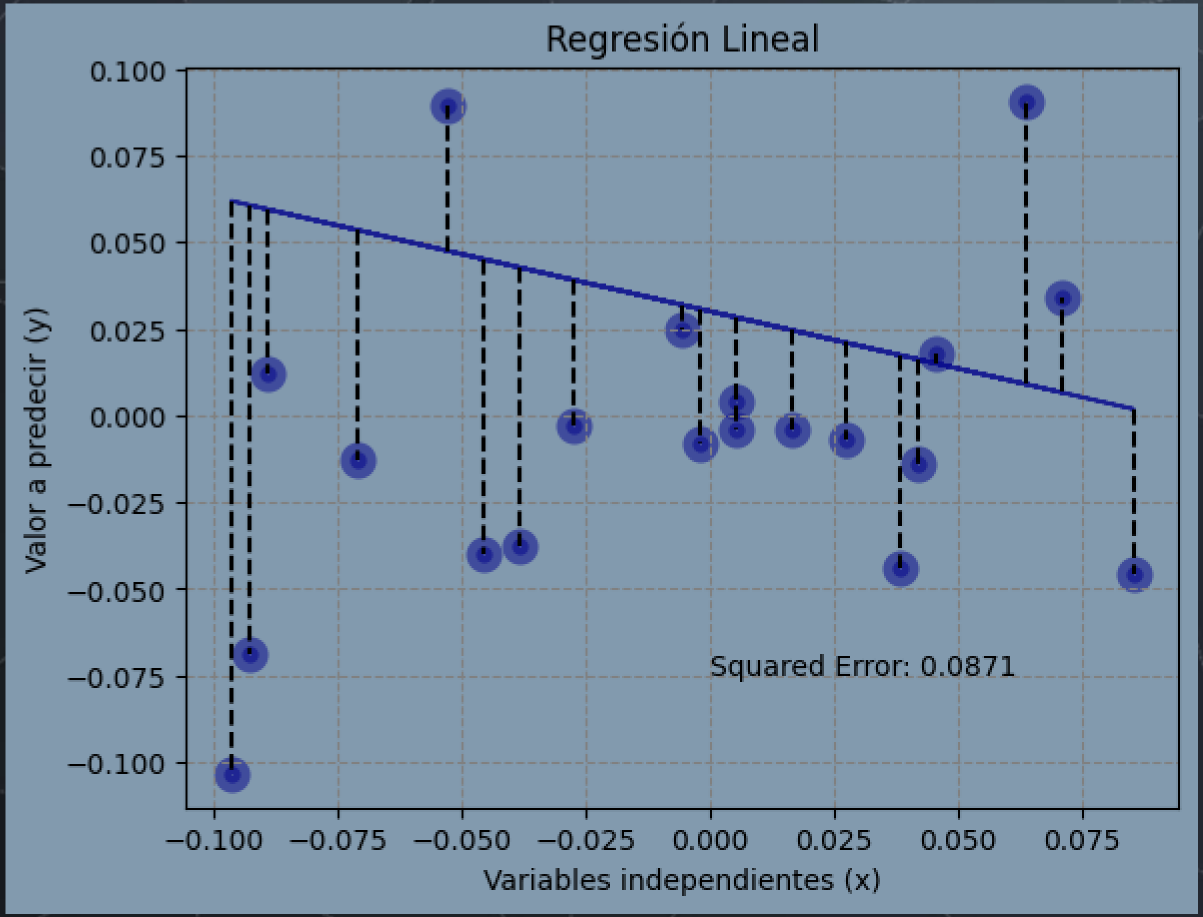
¿Pero cómo? Mínimos cuadrados.



TAE IA
Formación de Talento Altamente Especializado

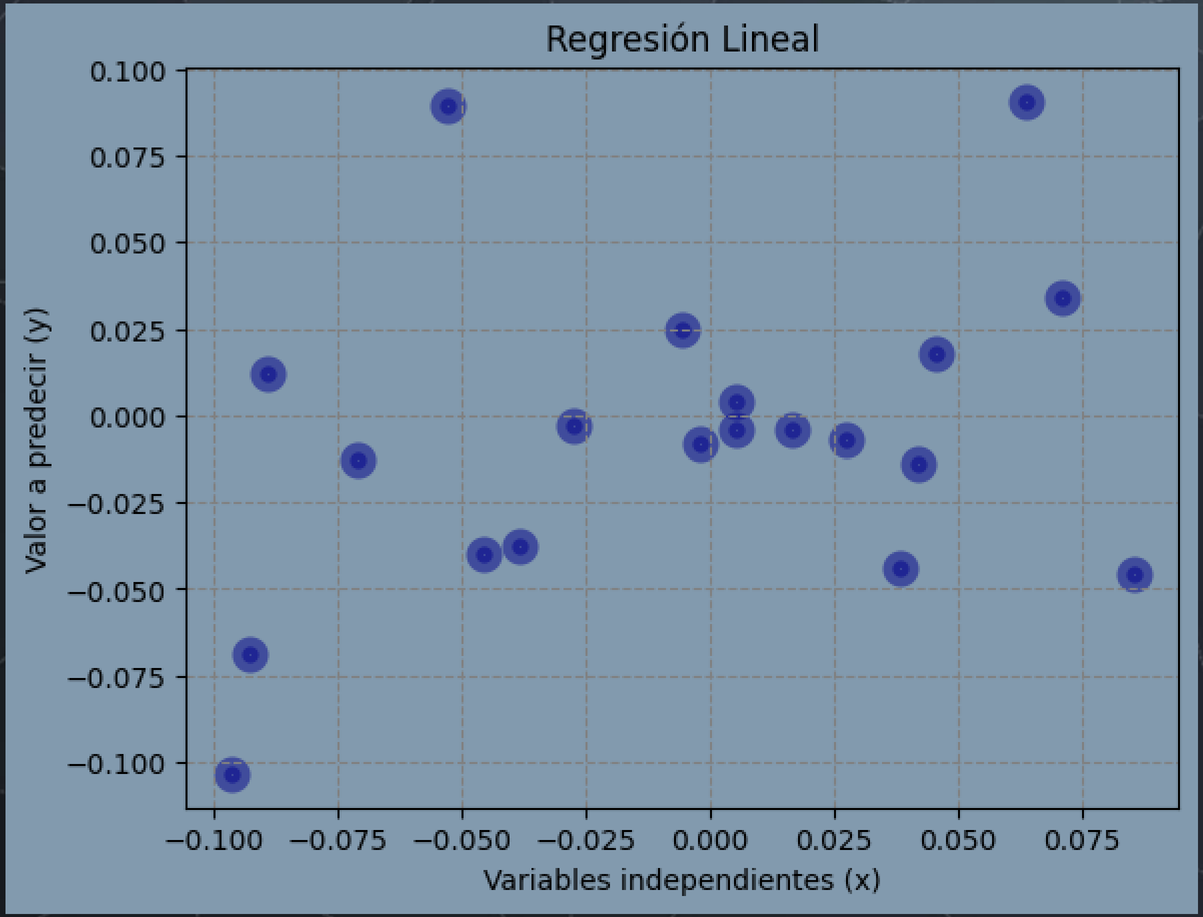
Regresión Lineal

¿Pero cómo? Mínimos cuadrados.



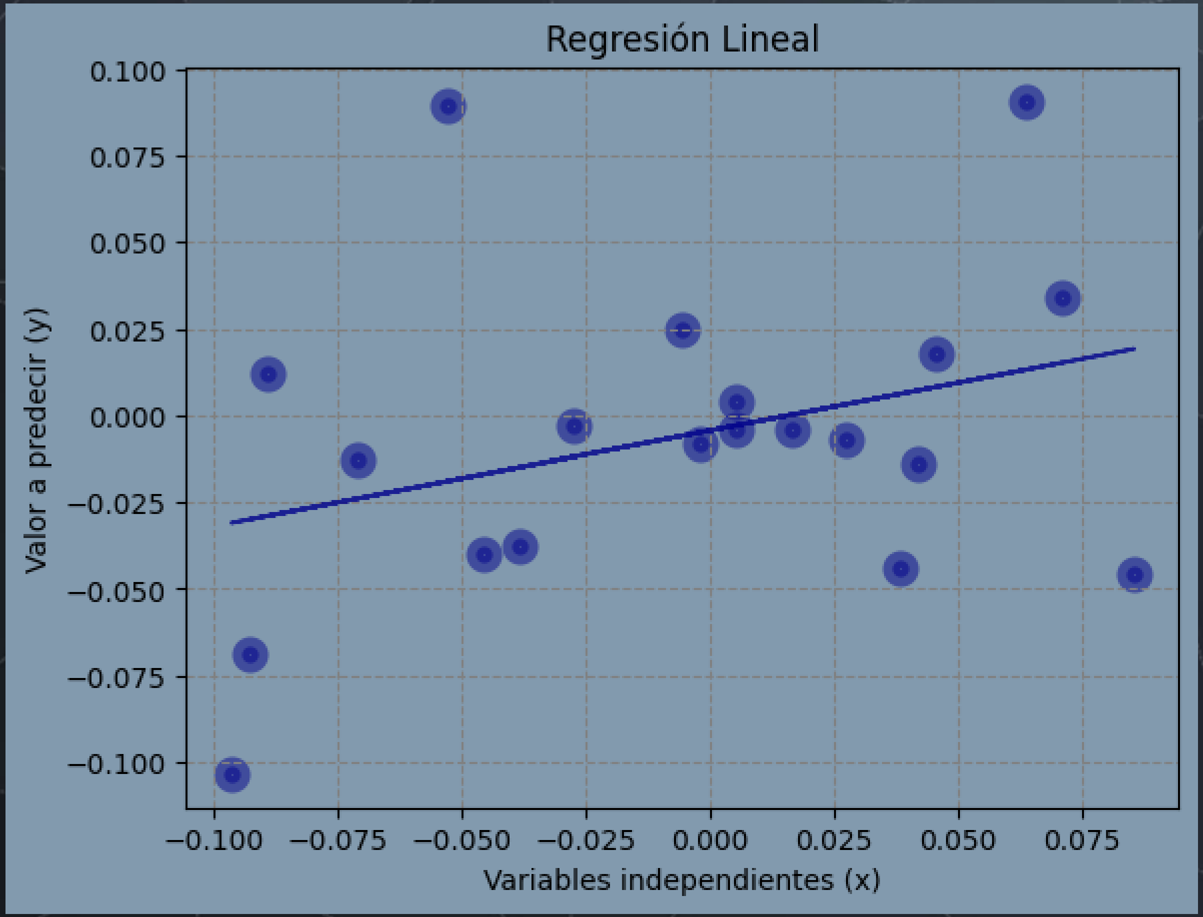
Regresión Lineal

¿Pero cómo? Mínimos cuadrados.



Regresión Lineal

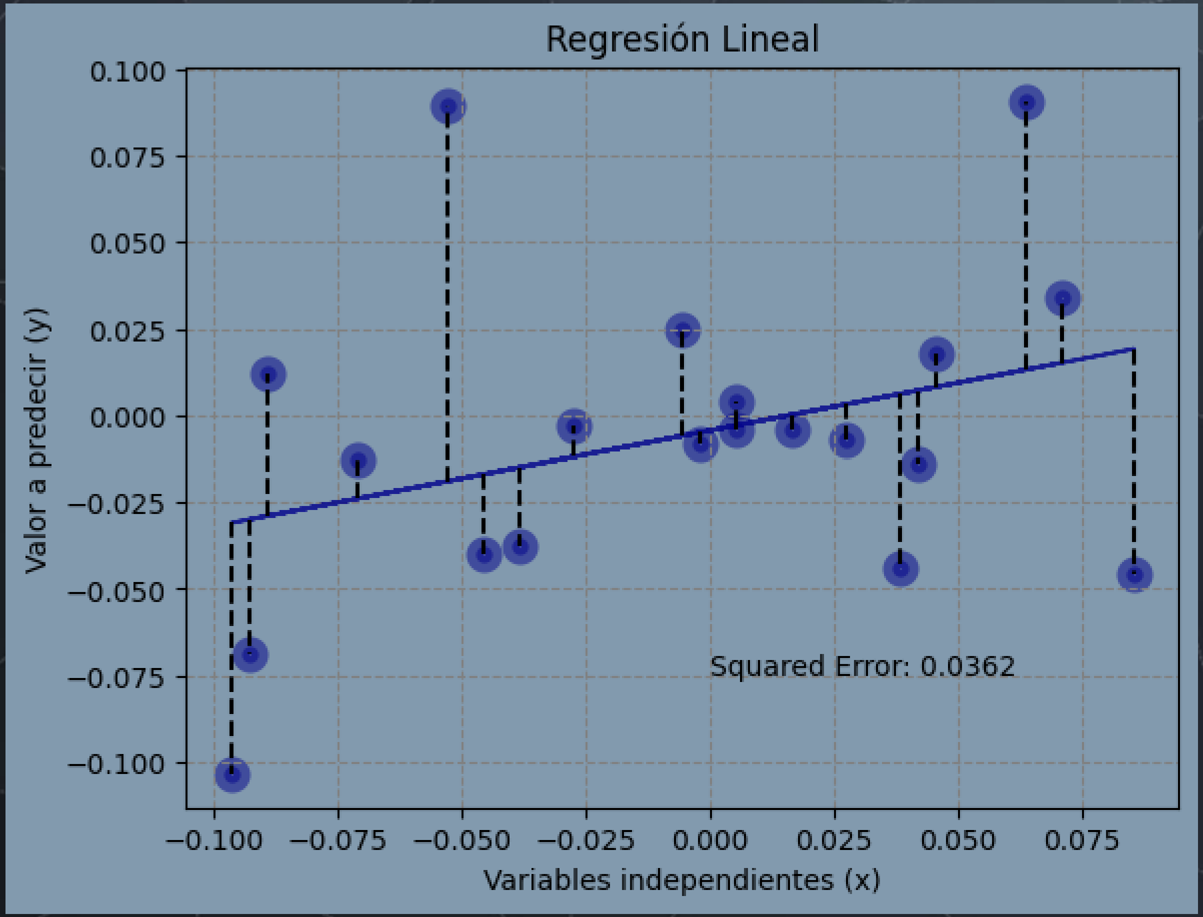
¿Pero cómo? Mínimos cuadrados.



TAE IA
Formación de Talento Altamente Especializado

Regresión Lineal

¿Pero cómo? Mínimos cuadrados.



Regresión Lineal

¿Pero cómo? Mínimos cuadrados.

$$min\left\{Suma(error^2)\right\}$$

$$Suma(error^2) = \sum (y_{real} - y_{modelo})^2$$

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum ((y_i - a_0 - a_1 x_i) x_i)$$

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

$$(n)a_0 + \left(\sum x_i\right) a_1 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i$$

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

Regresión Polinomial

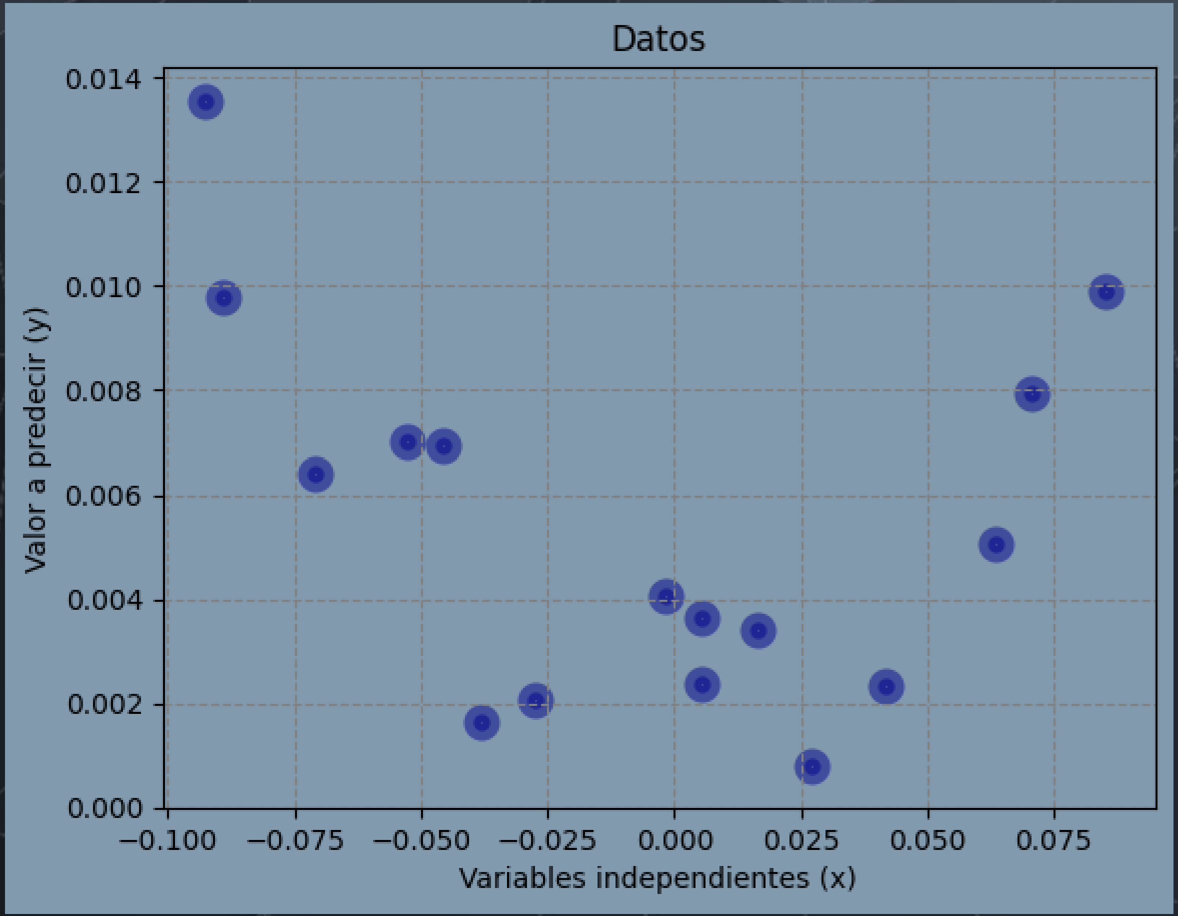
Generalización de la regresión lineal para una misma variable:

$$y = a_0 + a_1 X_1 + a_2 X_1^2 + \ \dots + a_g X_1^g$$

Esto le da una mayor “flexibilidad” al modelo para poder ajustarse a los datos.

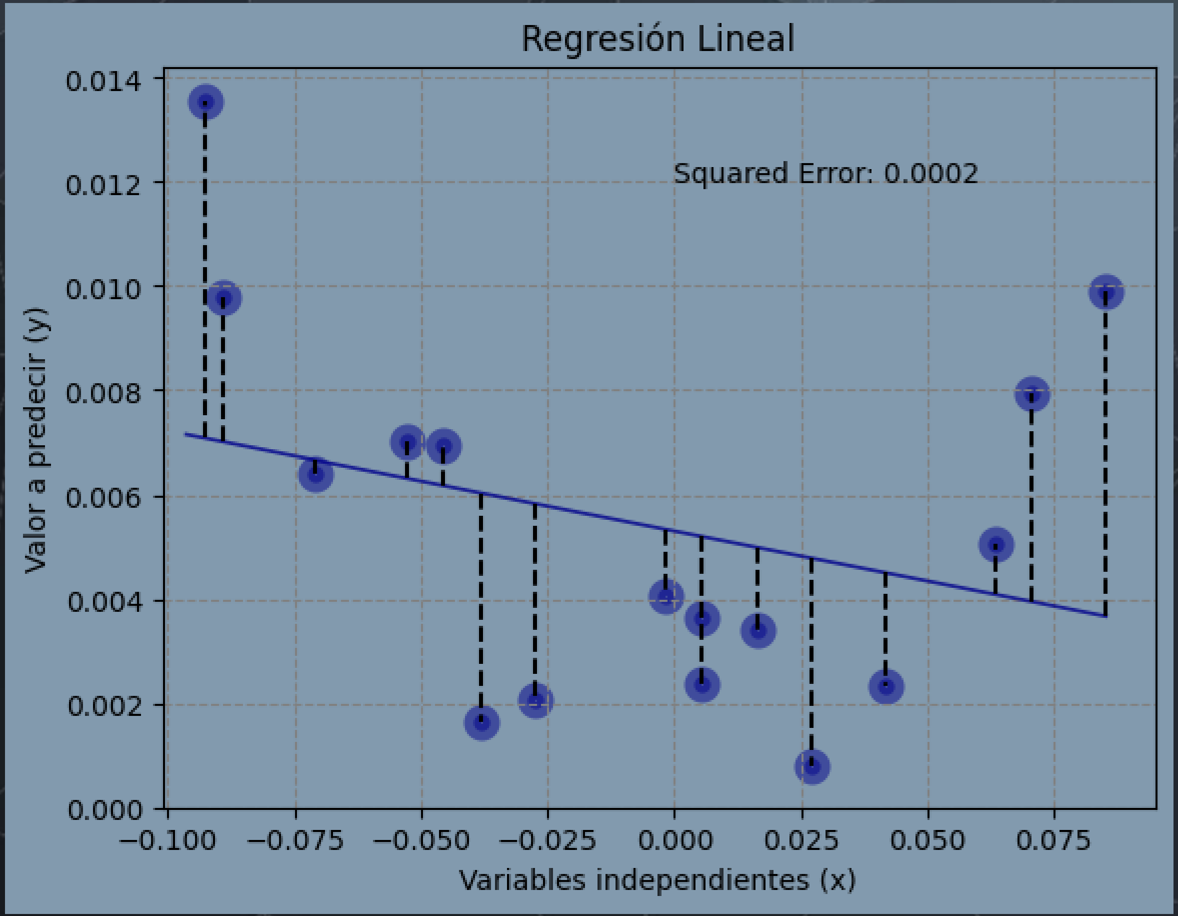
Regresión Polinomial

$$y = a_0 + a_1 X_1 + a_2 X_1^2 + \ldots + a_g X_1^g$$



Regresión Polinomial

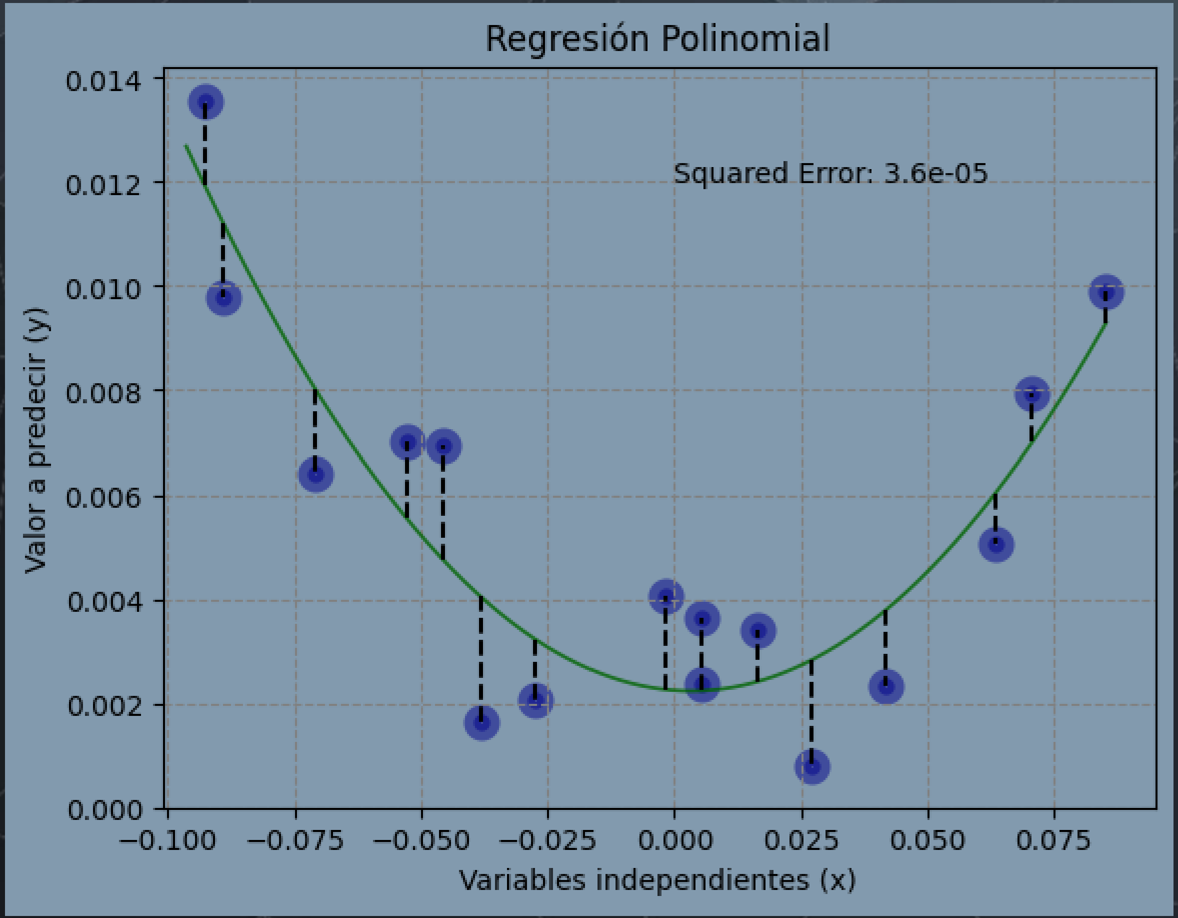
$$y = a_0 + a_1 X_1$$



TAE IA
Formación de Talento Altamente Especializado

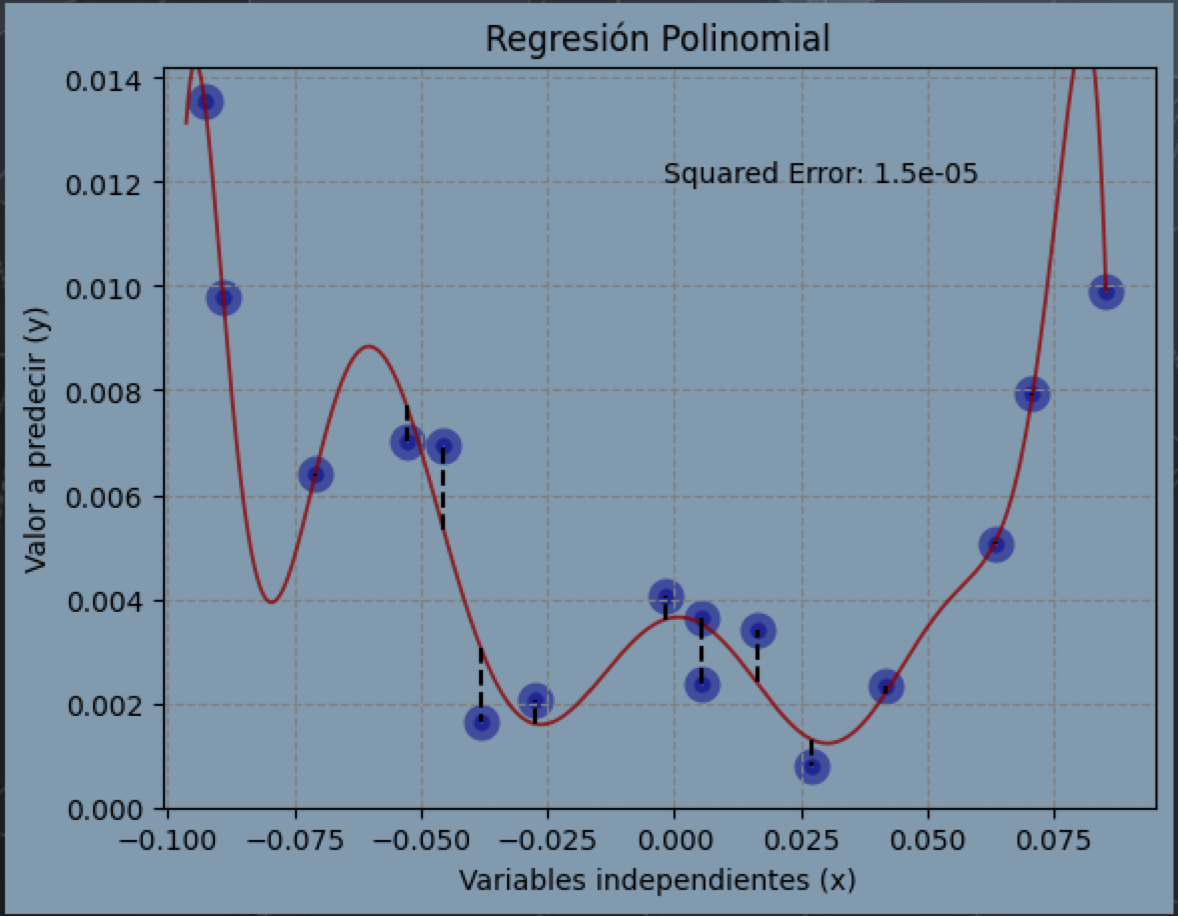
Regresión Polinomial

$$y = a_0 + a_1 X_1 + a_2 X_1^2$$



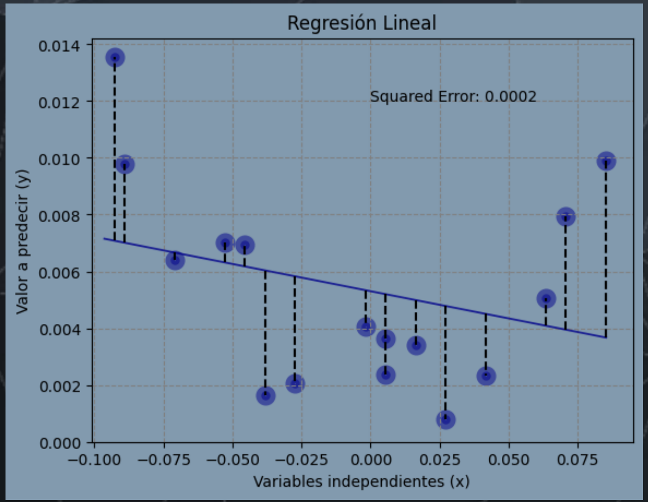
Regresión Polinomial

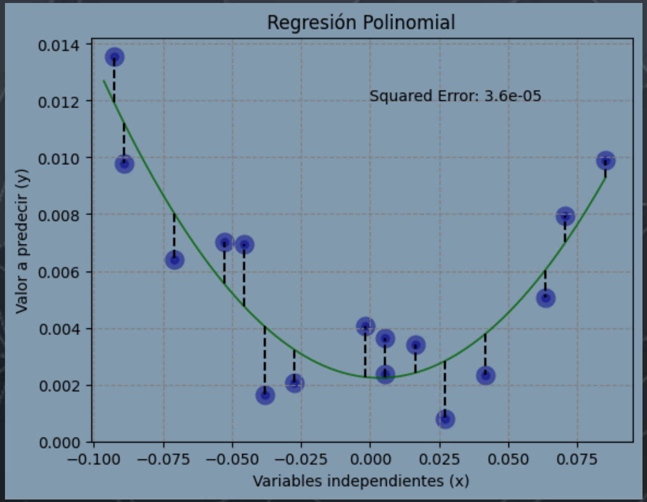
$$y = a_0 + a_1X_1 + a_2X_1^2 + \ ... + a_9X_1^9$$

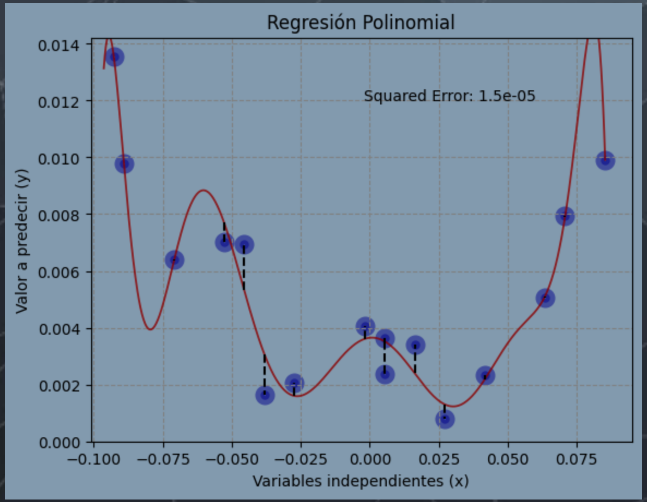


Regresión Polinomial

¿Qué modelo es el mejor?



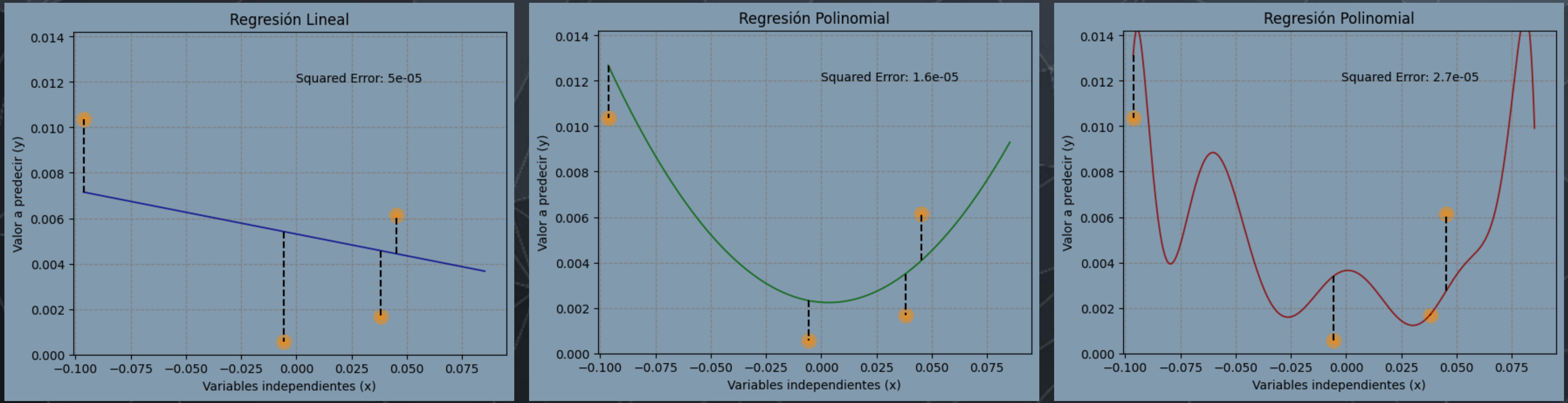




TAE IA
Formación de Talento Altamente Especializado

Regresión Polinomial

¿Qué modelo es el mejor?



TAE IA
Formación de Talento Altamente Especializado

Regresión Lineal

¡Vamos a probar!

Métricas de Regresión

Métricas de Regresión

Buscamos encontrar que tan cercanas las predicciones estuvieron de los valores que esperamos obtener

Error cuadrado promedio

Raíz del Error cuadrado promedio

Error absoluto promedio

Error absoluto porcentual promedio

Coeficiente de determinación

Métricas de Regresión

Buscamos encontrar que tan cercanas las predicciones estuvieron de los valores que esperamos obtener

- Error cuadrado promedio
- Raíz del Error cuadrado promedio
- Error absoluto promedio
- Error absoluto porcentual promedio
- Coeficiente de determinación

Métricas de Regresión

Error cuadrado promedio

- MSE (mean squared error)
- Determina el error (diferencia entre lo predicho y el valor real) al cuadrado (para eliminar negativos). De todo el conjunto de datos.
- Obtiene el promedio de esto.

$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2.$$

Métricas de Regresión

Error cuadrado promedio

- MSE (mean squared error)
- Determina el error (diferencia entre lo predicho y el valor real) al cuadrado (para eliminar negativos). De todo el conjunto de datos.
- Obtiene el promedio de esto.

$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2.$$

Raiz del MSE

- RMSE (root mean squared error)
- Solo se saca la raíz cuadrada del resultado de MSE:
- Esta métrica tiene las mismas unidades que la variable target (mayor interpretabilidad)

TAE IA
Formación de Talento Altamente Especializado

Métricas de Regresión

Error absoluto promedio

- MAE (mean absolute error)
- Similar al MSE, pero en lugar de obtener la diferencia al cuadrado, es solo el valor absoluto (L1-norm loss).

$$\text{MAE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \left| y_i - \hat{y}_i \right|.$$

Métricas de Regresión

Error absoluto promedio

- MAE (mean absolute error)
- Similar al MSE, pero en lugar de obtener la diferencia al cuadrado, es solo el valor absoluto (L1-norm loss).

$$\text{MAE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \left| y_i - \hat{y}_i \right|.$$

Error absoluto porcentual promedio

- MAPE (mean absolute percentage error)
- Métrica se enfoca en errores relativos, no es afectada por un cambio grande en la variable de respuesta.

$$\text{MAPE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \frac{\left| y_i - \hat{y}_i \right|}{\max(\epsilon, \left| y_i \right|)}$$

TAE IA
Formación de Talento Altamente Especializado

Métricas de Regresión

Coeficiente de determinación

- R2
- Representa la proporción de la varianza explicada por el modelo.
- Si explicamos la varianza entonces este score es 1 (modelo perfecto), si el score es 0 entonces es un modelo “constante”, es decir no contempla los features de entrada.
- Si el valor es menor a 0 nos indica que el modelo es peor que predecir solo el promedio de la variable de respuesta.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Referencias

- Scikit Learn (2024) Linear models. Recuperado el 25 de Julio del 2024 de: https://scikit-learn.org/stable/modules/linear_model.html
- Lewinston, E., (2023) A Comprehensive Overview of Regression Evaluation Metrics. NVIDIA Developper, recuperado el 25 de julio del 2024 de: https://developer.nvidia.com/blog/a-comprehensive-overview-of-regression-evaluation-metrics/
- Brownlee, J. (2021) Regression Metrics for Machine Learning. Recuperado el 25 de julio del 2024 de: https://machinelearningmastery.com/regression-metrics-for-machine-learning/
- Hastie, T.; Tibshirani, R. and Friedman, J. (2008) The elements of statistical learning Data Mining, Inference, and Prediction. Springer. Standford California.

Gracias por su atención

¿Dudas?

TAE

Formación de Talento Altamente Especializado

IA