

INFORME DE RESULTADOS DEL ANALISIS DEL DATASET BOSTON HOUSING.

El código desarrollado implementa un algoritmo de regresión lineal simple para predecir el valor promedio de las casas en Boston en función del número de habitaciones (RM) utilizando el conjunto de datos Boston Housing. Este conjunto de datos contiene información sobre diversas características de los hogares en la región de Boston, siendo una de ellas el número promedio de habitaciones. El objetivo es utilizar la regresión lineal para identificar cómo el número de habitaciones afecta el valor medio de las casas.

Descripción de la implementación:

- **Carga del dataset:** El conjunto de datos de Boston Housing se carga mediante la función `fetch_openml` de `sklearn.datasets`. Posteriormente, se selecciona la característica RM (número de habitaciones promedio por casa) como el predictor, mientras que la variable objetivo es MEDV (valor medio de las casas).
- **Inicialización de Parámetros:** Para implementar el algoritmo de regresión lineal, se definen dos parámetros clave: la pendiente m y el intercepto b , que son los que el algoritmo ajustará. Inicialmente, ambos valores son 0. También se establece una tasa de aprendizaje L de 0.01 y se fija el número de iteraciones (epochs) en 1000. La tasa de aprendizaje controla la velocidad con la que se ajustan los parámetros en cada iteración, mientras que el número de iteraciones define cuántas veces se actualizarán estos parámetros.
- **Algoritmo de Gradiente Descendente:** Se utiliza el algoritmo de gradiente descendente para minimizar el error entre los valores predichos por el modelo y los valores reales del conjunto de datos. En cada iteración, se calculan dos derivadas parciales: una respecto a la pendiente m y otra respecto al intercepto b . Estas derivadas se utilizan para actualizar ambos parámetros con el fin de reducir el error en cada paso. Después de 1000 iteraciones, el algoritmo debería haber ajustado los parámetros m y b lo suficiente para aproximar una línea de regresión adecuada.
- **Visualización:** Finalmente, se genera una visualización de los datos mediante un gráfico de dispersión. Los puntos amarillos representan los datos reales, mientras que la línea azul muestra la predicción del modelo ajustado tras aplicar el gradiente descendente.

Resultados:

El gráfico resultante muestra cómo la cantidad promedio de habitaciones influye en el valor medio de las casas. Se puede observar una tendencia positiva, donde a mayor número de habitaciones, el valor de las casas tiende a aumentar, lo que es coherente con las expectativas. La línea azul representa la predicción lineal, que sigue esta tendencia ascendente, aunque no todos los puntos están perfectamente alineados con la línea, lo que indica la presencia de cierta dispersión en los datos.

Adecuación del Algoritmo:

La regresión lineal simple es una herramienta adecuada para este tipo de problema cuando se desea observar la relación entre dos variables. En este caso, el número de habitaciones (RM) y el valor medio de las casas (MEDV). Sin embargo, el modelo tiene limitaciones ya que asume una relación lineal entre las variables. Dado que otros factores, como la ubicación y los ingresos, también influyen en el valor de las casas, este modelo podría no capturar todas las variaciones del problema real por lo que la precisión no es exacta.

Anexos.

