

Linear Regression

The California Housing Department wants to develop a **predictive model** to estimate the **median house value** in different regions of California based on various demographic and geographic factors.

You are provided with the dataset **housing.csv**, which contains housing-related information for different districts in California.

Your task is to **build a Linear Regression model from scratch using NumPy** to predict the **median_house_value** (target variable) and visualize the regression results using **Matplotlib**.

Dataset Details

Column	Description
longitude	Longitude coordinate of the district
latitude	Latitude coordinate of the district
housing_median_age	Median age of houses in the district
total_rooms	Total number of rooms within the district

Column	Description
total_bedrooms	Total number of bedrooms within the district
population	Total population of the district
households	Number of households in the district
median_income	Median income of the district (in tens of thousands of dollars)
median_house_value	Median value of houses (Target Variable)
ocean_proximity	Location category relative to the ocean (e.g., NEAR BAY, INLAND, etc.)

What to Do

- Implement **Linear Regression from scratch using NumPy**.
- Load and prepare the dataset
- Train your Linear Regression model and **visualize the regression fit** or residuals using **Matplotlib**.
- Output the final performance score you obtain.

Logistic Regression

The **National Health Research Institute** aims to develop a predictive model to determine whether a patient is likely to have **diabetes** based on their demographic, lifestyle, and health parameters.

You are provided with the dataset **diabetes.csv**, which contains patient health records, including blood glucose levels, BMI, age, and other clinical indicators.

Your task is to **implement Logistic Regression from scratch using NumPy** to predict the likelihood of a patient having diabetes. You will train, visualize, and evaluate your model using **Matplotlib** for clear insights.

Dataset Details	
Column	Description
year	Year of data collection
gender	Gender of the patient (Male/Female)
age	Age of the patient

Column	Description
location	State or region where the patient resides
race:AfricanAmerican, race:Asian, race:Caucasian, race:Hispanic, race:Other	Encoded race features
hypertension	1 if the patient has hypertension, 0 otherwise
heart_disease	1 if the patient has heart disease, 0 otherwise
smoking_history	Smoking status (never, current, former, etc.)
bmi	Body Mass Index (BMI)
hbA1c_level	Average blood sugar level over the past 3 months
blood_glucose_level	Measured blood glucose level

Column	Description
diabetes	Target variable (1 = diabetic, 0 = non-diabetic)

- **What to Do**

- Implement **Logistic Regression from scratch using NumPy** (do not use sklearn or other ML libraries).
- Start by **loading and preprocessing the dataset** .
- Train your Logistic Regression model and visualize important aspects (such as **loss over iterations** or feature relationships) using **Matplotlib**.
- Output the **accuracy** your model achieves on the test set.