Linear Regression Model

Linear regression is a statistical method that is used for modeling the relationship between a dependent variable and one or more independent variables which is done by fitting a linear equation to observed data. For this method it is assumed that there is a linear relationship between the independent variables and the dependent variable.

The Simple linear regression

The equation of a simple linear regression with one independent variable is represented as follows:

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon$$

Where:

y is the dependent variable.

x is the independent variable.

β_0 is the y-intercept of the line, representing the value of y when x = 0.

β_1 is the slope of the line, representing the change in y for a one-unit change in x.

ε is the error term, representing the difference between the observed and predicted values of y.

The goal of linear regression is finding the best-fitting line through the data points, so the sum of the squared differences between the observed and predicted values is minimized.

Practical Example:

Let's consider a practical example for predicting apartment prices based on their area. We have a dataset containing information about the area of houses and their corresponding prices. The dataset is placed in a csv file named apartments.csv and we read the dataset from it.

The dataset from the csv file is also given below:

Size (m^2)	Price ($)
25.3	50000
26.8	51000
30.5	53000
35.2	57000
40.6	61000
45.9	65000
50.4	69000
55.1	73000
60.7	77000
65.2	81000
70.9	85000
75.5	89000
80.3	93000
85.6	97000
90.1	101000
95.8	105000
100.4	109000
105.2	113000
110.7	117000
115.3	121000
120	125000
26.7	53000
31.2	57000
36.5	61000
41.8	65000
46.4	69000
51	73000
56.7	77000
61.3	81000
66.8	85000

Python Code

```python
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression

# Read data from CSV file
data = pd.read_csv("apartments.csv")

# Separate independent and dependent variables
X = data.iloc[:, :-1]  # Features (size of flats)
y = data.iloc[:, -1]   # Target (prices)

# Create a linear regression model
model = LinearRegression()

# Fit the model to the data
model.fit(X, y)

# Predict prices for new data points
new_X = np.array([[28.6], [75.2], [230.9]])
predicted_prices = model.predict(new_X)

# Output the predicted prices
for i, price in enumerate(predicted_prices):
    print(f"Predicted price for {new_X[i][0]} m²: ${price:.2f}")
```

We trained the model and then used it to calculate prices for the apartment sizes 28.6, 75.2, and 230.9 meters squared. Here are the results:

```
Predicted price for 28.6 m²: $53330.21
Predicted price for 75.2 m²: $89557.54
Predicted price for 230.9 m²: $210600.34
```