```
# Lab 5: Multivariable Regression
# 0. Installation and Import Libraries

# If not previously installed, install ucimlrepo
!pip install ucimlrepo

from ucimlrepo import fetch_ucirepo
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
```

```
Collecting ucimlrepo
  Downloading ucimlrepo-0.0.7-py3-none-any.whl.metadata (5.5 kB)
Requirement already satisfied: pandas>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from ucimlrepo) (2.2.2)
Requirement already satisfied: certifi>=2020.12.5 in /usr/local/lib/python3.12/dist-packages (from ucimlrepo) (2025.1
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->ucimlrep
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0-
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->ucimlrepo
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.0.0->ucimlre
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pand
Downloading ucimlrepo-0.0.7-py3-none-any.whl (8.0 kB)
Installing collected packages: ucimlrepo
Successfully installed ucimlrepo-0.0.7
```

```
# Lab 5: Multivariable Regression
# 1. Load Dataset and Inspect

computer_hardware = fetch_ucirepo(id=29)

# Features and target extraction
X = computer_hardware.data.features
# Extract 'ERP' as the target variable and drop it from features
y = X['ERP']
X = X.drop('ERP', axis=1)


# Metadata and variable info
print("Metadata:\n", computer_hardware.metadata)
print("\nVariable Information:\n", computer_hardware.variables)

# Inspect top rows
print("Sample features:\n", X.head())
print("Sample targets:\n", y.head())
```

Toggle Gemini

```
Metadata:
 {'uci_id': 29, 'name': 'Computer Hardware', 'repository_url': 'https://archive.ics.uci.edu/dataset/29/computer+hardw

Variable Information:
          name     role         type demographic  \
0  VendorName  Feature  Categorical        None
1   ModelName  Feature  Categorical        None
2        MYCT  Feature      Integer        None
3        MMIN  Feature      Integer        None
4        MMAX  Feature      Integer        None
5        CACH  Feature      Integer        None
6       CHMIN  Feature      Integer        None
7       CHMAX  Feature      Integer        None
8         PRP  Feature      Integer        None
9         ERP  Feature      Integer        None

                                         description        units  \
0  (adviser, amdahl,apollo, basf, bti, burroughs,...         None
1                                  many unique symbols         None
2                                   machine cycle time  nanoseconds
3                                  minimum main memory    kilobytes
4                                  maximum main memory    kilobytes
5                                         cache memory    kilobytes
6                                     minimum channels        units
7                                     maximum channels        units
8                       published relative performance         None
9  estimated relative performance from the origin...         None

  missing_values
0             no
1             no
2             no
3             no
4             no
5             no
6             no
```

```
7             no
8             no
9             no
Sample features:
   VendorName ModelName  MYCT  MMIN   MMAX  CACH  CHMIN  CHMAX  PRP
0    adviser     32/60   125   256   6000   256     16    128  198
1     amdahl    470v/7    29  8000  32000    32      8     32  269
2     amdahl   470v/7a    29  8000  32000    32      8     32  220
3     amdahl   470v/7b    29  8000  32000    32      8     32  172
4     amdahl   470v/7c    29  8000  16000    32      8     16  132
Sample targets:
 0    199
1    253
2    253
3    253
4    132
Name: ERP, dtype: int64
```

```
# Lab 5: Multivariable Regression
# 2. Train-Test Split

# Select only numerical features for splitting
X_numeric = X.select_dtypes(include=np.number)

X_train, X_test, y_train, y_test = train_test_split(
    X_numeric, y.values.ravel(), test_size=0.2, random_state=42
)

print("Training set size:", X_train.shape)
print("Testing set size:", X_test.shape)
```

```
Training set size: (167, 7)
Testing set size: (42, 7)
```

```
# Lab 5: Multivariable Regression
# 3. Model Fitting and Coefficients

# Create and fit the linear regression model
lr = LinearRegression()
lr.fit(X_train, y_train)

# Print intercept and coefficients
print("Intercept (b0):", lr.intercept_)
print("Coefficients (b1, b2, ...):", lr.coef_)
```

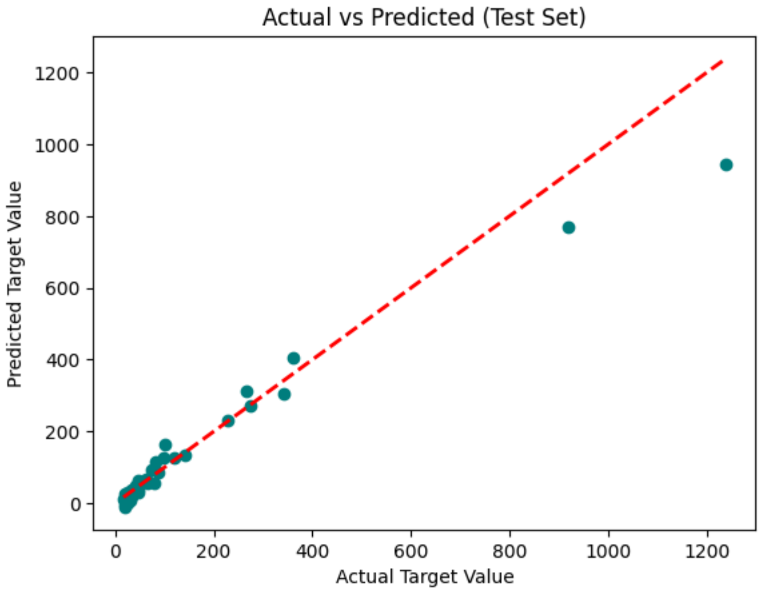
```
Intercept (b0): -29.84146538350059
Coefficients (b1, b2, ...): [ 0.03216256  0.00360426  0.00408698  0.32243938 -0.23783053  0.23001715
  0.47875042]
```

Toggle Gemini

```
# Lab 5: Multivariable Regression
# 4. Prediction, Output, and Evaluation

# Predict on the test set
y_pred = lr.predict(X_test)

# Output: Mean squared error and R2 score
print("Mean Squared Error:", mean_squared_error(y_test, y_pred))
print("R2 Score:", r2_score(y_test, y_pred))
```

```
Mean Squared Error: 3007.8898321639304
R2 Score: 0.9440465034138787
```

```
# Lab 5: Multivariable Regression
# 5. Visualization (Actual vs Predicted)

plt.scatter(y_test, y_pred, color='teal')
plt.xlabel('Actual Target Value')
plt.ylabel('Predicted Target Value')
plt.title('Actual vs Predicted (Test Set)')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'r--', lw=2)
plt.show()
```



```
Start coding or generate with AI.
```

Toggle Gemini