Homework 6: Predicting Movie Revenue

Project Overview

This report presents a linear regression approach to predicting movie revenue using Python and scikit-learn. The model uses production budget, genre popularity, and cast influence as predictors. The goal is to estimate box office revenue based on these factors.

Implementation Highlights

- **Importing Libraries:** Libraries like pandas, numpy, and scikit-learn were imported to work with data and perform regression.

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

- **Reading the Dataset:** The Excel file was read into a pandas DataFrame.

```
excelData = pd.read_excel("HW 6 Data.xlsx")
```

- **Selecting Features and Target:** Three input features were used to predict the revenue.

```
features = excelData[['ProductionBudget',
      'GenrePopularity', 'CastStarPower']]
boxOfficeTarget = excelData['BoxOfficeRevenue']
```

- **Training the Model:** A linear regression model was created and trained on the selected features.

```
model = LinearRegression()
model.fit(features, boxOfficeTarget)
```

- **Printing Model Summary:** The intercept and weights (coefficients) for each feature were displayed.

```
print("\n Linear Model Summary")
print(f"→ Intercept:  ${model.intercept_:.2f} million")
for label, weight in zip(features.columns, model.coef_):
  print(f"→ Weight for '{label}':  {weight:.2f}")
```

- **Prediction Example:** A prediction was made using new movie inputs.

```
testInput = pd.DataFrame([[120, 8, 6]], columns=features.columns)
estimatedRevenue = model.predict(testInput)[0]
print("\n Prediction Result:")
print("Given the values - Budget:  $120M, Genre Popularity:  8,
Cast Star Power:  6")
print(f" Predicted Box Office Revenue:  ${estimatedRevenue:.2f}
million")
```

Equation Derived from Model

$$\text{Revenue} = -103.86 + 3.60 \times \text{Budget} + 9.80 \times \text{Genre} + 13.56 \times \text{Cast}$$

Prediction Example

Feature	Value
Production Budget	120 million
Genre Popularity	8
Cast Star Power	6

$$\text{Revenue} = -103.86 + 432 + 78.4 + 81.36 = \textbf{487.82 million}$$

Interpretation: Based on the trained model, a movie with a production budget of 120 million, genre popularity of 8, and cast star power of 6 is predicted to earn approximately **487.82 million** at the box office. This estimate reflects the cumulative influence of all three input factors, with cast star power contributing the most significant weight among the features.

Submission References

- **Google Colab:** https://colab.research.google.com/drive/1LRnPvhHdYwXlXGBUZTg90_Cegz00j_v6

Visual Output

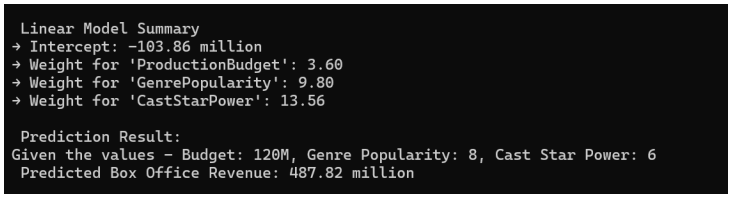

Figure 1: Model Prediction Output