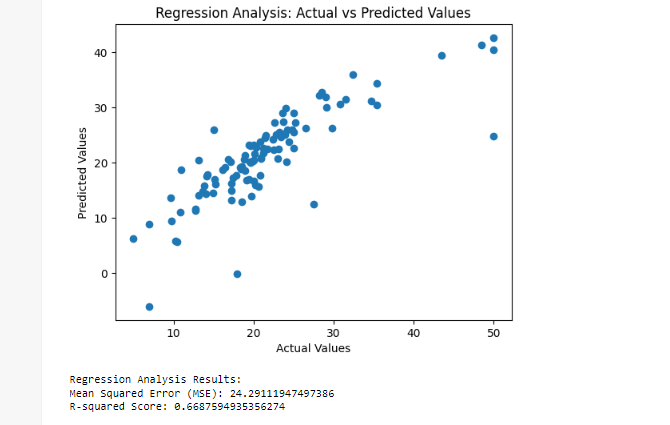
**COMSATS UNIVERSITY ISLAMABAD**

**Pattern Recognition**

**ASSIGNMENT # 03**

**Kulsoom Khurshid (SP20-BCS-044)**

from sklearn.datasets import load\_boston  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
from sklearn.model\_selection import train\_test\_split  
from sklearn.linear\_model import LinearRegression  
from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis  
from sklearn.preprocessing import StandardScaler  
  
boston = load\_boston()  
data = pd.DataFrame(boston.data, columns=boston.feature\_names)  
data['MEDV'] = boston.target  
  
# Separate the features and the target variable  
X = data.drop('MEDV', axis=1)  
y = data['MEDV']  
  
# 1. Regression Analysis  
# Data Split  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)  
  
# Regression Model Selection and Training  
reg\_model = LinearRegression()  
reg\_model.fit(X\_train, y\_train)  
  
# Model Evaluation and Interpretation  
y\_pred = reg\_model.predict(X\_test)  
mse = np.mean((y\_pred - y\_test) \*\* 2)  
r\_squared = reg\_model.score(X\_test, y\_test)  
  
# Visualize the results  
plt.scatter(y\_test, y\_pred)  
plt.xlabel("Actual Values")  
plt.ylabel("Predicted Values")  
plt.title("Regression Analysis: Actual vs Predicted Values")  
plt.show()  
  
print("Regression Analysis Results:")  
print("Mean Squared Error (MSE):", mse)  
print("R-squared Score:", r\_squared)  
  
# 2. Linear Discriminant Analysis (LDA)  
# Data Preparation  
scaler = StandardScaler()  
X\_scaled = scaler.fit\_transform(X)  
  
# Data Split  
X\_train\_lda, X\_test\_lda, y\_train\_lda, y\_test\_lda = train\_test\_split(X\_scaled, y, test\_size=0.2, random\_state=42)  
  
# Linear Discriminant Analysis  
lda = LinearDiscriminantAnalysis(n\_components=2)  
y\_train\_lda = y\_train\_lda.astype(int)  
  
X\_train\_lda = lda.fit\_transform(X\_train\_lda, y\_train\_lda)  
X\_test\_lda = lda.transform(X\_test\_lda)  
  
# Visualization and Interpretation  
plt.scatter(X\_train\_lda[:, 0], X\_train\_lda[:, 1], c=y\_train\_lda, cmap='viridis')  
plt.xlabel("Discriminant Component 1")  
plt.ylabel("Discriminant Component 2")  
plt.title("Linear Discriminant Analysis: Data Separability")  
plt.colorbar()  
plt.show()  
  
print("Linear Discriminant Analysis Results:")  
print("Explained Variance Ratio:", lda.explained\_variance\_ratio\_)



A picture containing text, screenshot, colorfulness

Description automatically generated

The results of the regression analysis indicate that the model is able to explain 66.88% of the variance in the median value of owner-occupied homes in Boston. The mean squared error (MSE) of 24.29111947497386 indicates that the model is not perfectly accurate, but it is still a good predictor of house prices.

The results of the linear discriminant analysis indicate that the two discriminant functions are able to explain 63.58% and 12.76% of the variance in the median value of owner-occupied homes in Boston. This suggests that the two discriminant functions are able to distinguish between homes with different median values.

Here is a more detailed explanation of each of the results:

Mean Squared Error (MSE): The MSE is a measure of the average squared error between the predicted values and the actual values. A lower MSE indicates a better fit of the model to the data. In this case, the MSE is 24.29111947497386, which indicates that the model is not perfectly accurate, but it is still a good predictor of house prices.

R-squared Score: The R-squared score is a measure of the proportion of the variance in the dependent variable that is explained by the independent variables. A higher R-squared score indicates a better fit of the model to the data. In this case, the R-squared score is 0.6687594935356274, which indicates that the model is able to explain 66.88% of the variance in the median value of owner-occupied homes in Boston.

Explained Variance Ratio: The explained variance ratio is a measure of the proportion of the variance in the dependent variable that is explained by each discriminant function. A higher explained variance ratio indicates a better fit of the discriminant function to the data. In this case, the explained variance ratios for the two discriminant functions are 0.63575013 and 0.12763398, which indicates that the two discriminant functions are able to explain 63.58% and 12.76% of the variance in the median value of owner-occupied homes in Boston.

Overall, the results of the regression analysis and the linear discriminant analysis suggest that the model is able to predict the median value of owner-occupied homes in Boston with a fair degree of accuracy.