**Project 1- Iris Data Classification**

Latheesh Mangeri  
1002116009  
2242-CSE-6363-005

**Problem:**

Classification on the Iris dataset using linear regression and cross-validation. The goal is to build a model that can accurately predict the species of an Iris flower based on its sepal and petal measurements.

**Data:**

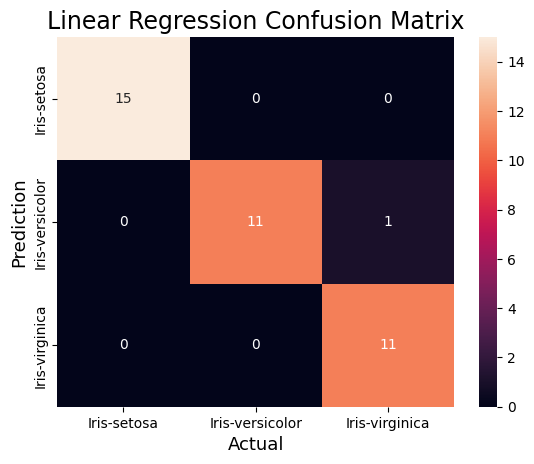
The Iris dataset contains 150 data points, each representing an Iris flower. Each data point has four features: sepal length, sepal width, petal length, and petal width. The label is the species of the flower, which can be Iris setosa, Iris versicolor, or Iris virginica. You can download the data from the provided URL.

**Methodology:**

1. **Preprocessing:**
   1. Load the data into your preferred programming language (e.g., Python, R).
   2. Split the data into features and labels.
   3. Consider encoding the categorical label (species) into numerical values.
   4. Normalize or standardize the features if necessary.
2. **Model Training:**
   1. Implemented a linear regression model.
   2. Use k-fold cross-validation (k=10) to train and evaluate the model.
   3. Within each fold:
      1. Split the data into training and testing sets by 75-25.
      2. Trained the linear regression model on the training set.
      3. Evaluated the model performance on the testing set using metrics like accuracy, precision, recall, and F1-score.
3. **Classification:**
   1. Train the final model on the entire dataset without cross-validation.
   2. Use the trained model to predict the species of new Iris flowers based on their measurements.

**Results:**

* **Confusion Matrix:**

****

* **Classification Report**

**A screenshot of a computer screen

Description automatically generated**

**Conclusion:**

We used Linear Regression and secured an accuracy of 0.9736.

A graph with black text

Description automatically generated with medium confidence