CSE-6363-105-MACHINE LEARNING

Project-1 Report

Name: Kanthreegala Padmakar Koushik

UTA id: 1002049853

Data Analysis:

Dataset contains data of Iris flower with dimensions that are Petal Length, Petal Width, Sepal Length, Sepal Width, Species Type and are represented by x1, x2, x3, x4, class respectively. Example of data is shown below.

	x1	x2	x3	x4	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa

A visualization representation of input features corelation is done using seaborn library and shown in the following figure.

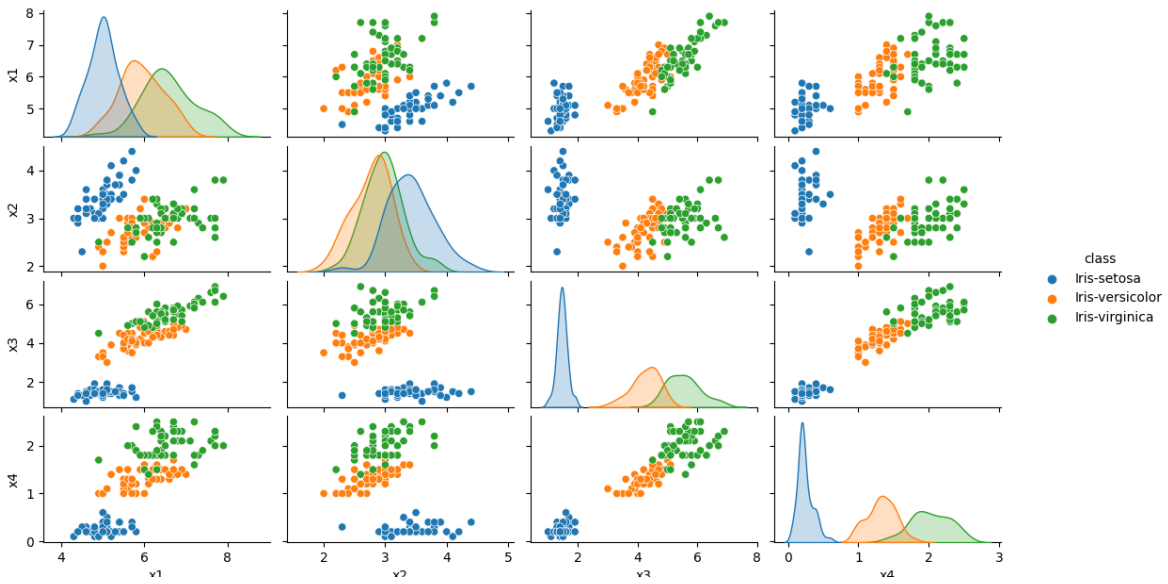

Figure.1 – Data visualization.

As we can see in the data visualization most data points are linearly separable and Iris-setosa class is most discreet class and given problem is to linearly regression so following mapping to the classes is proposed 'Iris-setosa' : 0, 'Iris-versicolor' : 2, 'Iris-virginica' : 3.

1. Training a model via linear regression:

Implementation is done using linear_regression class with fit() and predict().

Fit()- to train model using least square estimator.

Predict()- to predict output on unknown values.

2. Using the trained model to do the classification:

Out of 150 samples 135 samples are used for training and 15 samples are used for testing the model accuracy and the result for classification of model is :**100%**. To further verify the approach K-folds cross-validation is used.

Weight matrix for classification including bias is

Coefficients of linear model:

[1.0080465, -0.16229207, -0.28712107, 0.4524179, 0.64372899]

3. Using cross-validation.

Mean accuracy on entire data for 1-10 folds is shown below:

Mean Accuracy for 1-folds: 97.33%

Mean Accuracy for 2-folds: 97.33%

Mean Accuracy for 3-folds: 100.00%

Mean Accuracy for 4-folds: 100.00%

Mean Accuracy for 5-folds: 96.67%

Mean Accuracy for 6-folds: 100.00%

Mean Accuracy for 7-folds: 90.91%

Mean Accuracy for 8-folds: 100.00%

Mean Accuracy for 9-folds: 100.00%

Mean Accuracy for 10-folds: 93.33%