**Experiment No:1**

**Title:** Using only NumPy, design a simple neural network to classify the Iris flowers into three species based on sepal length, sepal width, petal length, and petal width.

After completion of this experiment students will have learned to train our own supervised machine learning model using Iris flower classification. Through this experiment they will learn about Machine Learning, Data Analysis, Data Visualization, Model Creation etc.

**Aim:** The Iris flower classification is to predict flowers based on their specific features.

**Theory:**

**Prerequisites:**

**NUMPY/SCIPY:-**

NumPy is a Python library, which stands for ‘Numerical Python’. It is the core library for scientific computing, which contains a powerful n-dimensional array object, provide tools for integrating C, C++ etc.

* NumPy array can also be used as an efficient multi-dimensional container for generic .data.
* The nparray (NumPy Array) is a multidimensional array used to store values of same datatype. These arrays are indexed just like Sequences, starting with zero.
* The nparrays are better than regular arrays in terms of faster computation and ease of manipulation.
* In different algorithms of Machine Learning like K-means Clustering, Random Forest etc. we have to store the values in an array. So, instead of using a regular array, nparray helps us to manipulate and execute easily.

**SCIKIT:-**

The functionality that scikit-learn provides include:

* **Regression**, including Linear and Logistic Regression
* **Classification**, including K-Nearest Neighbors
* **Clustering**, including K-Means and K-Means++

# Model selection

* **Preprocessing**, including Min-Max Normalization.

**PANDAS:-**

* Merging and Joining Data Sets.
* Reshaping & pivoting Data Sets.
* Inserting & deleting columns in Data Structure.
* Aligning data & dealing with missing data.
* Iterating over a Data set.
* Analyzing Time Series.
* Filtering Data around a condition.
* Arranging Data in an ascending & descending.
* Reading from files with CSV, TXT, XLSX, and other formats.
* Manipulating Data using integrated indexing for DataFrame objects.
* Generating Data range, date shifting, lagging, converting frequency, and other other Time Series functionality.
* Subsetting fancy indexing, & label based slicing Data Sets that are large in size.
* Performing split apply combine on Data Sets using the group by engine. With Python Pandas, it is easier to clean & wrangle with your Data. features of Pandas make it a great choice for Data Science and Analysis.

**MATPLOTLIB:-**

* Matplotlib is a visualization library in Python for 2D plots of arrays. It consists of several plots like line, bar, scatter, histogram etc.
* Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy [stack.It](http://stack.it/) can also be used with graphics toolkits like PyQt and wxPython.
* One of the advantages of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals.

**Steps to Classify Iris Flower:**

A purple flower with green leaves

Description automatically generated

1. Load the data
2. Analyze and visualizes the dataset
3. Model training
4. Model evaluation
5. Testing the data

Step 1: Load the data

A screenshot of a computer code

Description automatically generated

First, we have imported necessary packages for the project.

* Numpy will be used for any computational operations.
* We’ll use Matplotlib and seaborn for data visualization.
* Pandas helps to load data from various sources like local storage, database, excel file, CSV file, etc.

A white background with black text

Description automatically generated

* Next, we load the data using pd.read\_csv() and set the column name as per the iris data information.
* Pd.read\_csv reads CSV files. CSV stands for comma separated value.
* Df.head() only shows the first 5 rows from the data set table.

A screenshot of a computer

Description automatically generated

* All numerical values are in Centimeters.

Step 2. Analyze and visualizes the dataset

Here is some information about the dataset

A screenshot of a computer

Description automatically generated

From this description, we can see all the descriptions about the data, like average length and width, minimum value, maximum value, the 25%, 50%, and 75% distribution value, etc.

Let’s visualize the dataset.

A close up of a text

Description automatically generated

* To visualize the whole dataset we used the seaborn pair plot method. It plots the whole dataset’s information.

A group of graphs showing different sizes of data

Description automatically generated

* From this visualization, we can tell that iris-setosa is well separated from the other two flowers.
* And iris virginica is the longest flower and iris setosa is the shortest.

Now let’s plot the average of each feature of each class.

A white background with text

Description automatically generated

* Here we separated the features from the target value.

A screenshot of a computer code

Description automatically generated

* Np. average calculates the average from an array.
* Here we used two for loops inside a list. This is known as list comprehension.
* List comprehension helps to reduce the number of lines of code.
* The Y Data is a 1D array, but we have 4 features for every 3 classes. So we reshaped Y Data to a (4, 3) shaped array.
* Then we change the axis of the reshaped matrix.

A screenshot of a computer code

Description automatically generated

* We used matplotlib to show the averages in a bar plot.

A graph with different colored bars

Description automatically generated

* Here we can clearly see the verginica is the longest and setosa is the shortest flower.

3. Model training

A screenshot of a test

Description automatically generated

* Using a train test split we split the whole data into training and testing datasets. Later we’ll use the testing dataset to check the accuracy of the model.

A screenshot of a computer

Description automatically generated

* Here we imported a support vector classifier from the scikit-learn support vector machine.
* Then, we created an object and named it svn.
* After that, we feed the training dataset into the algorithm by using the svn.fit() method.

Step 4: Module Evaluation:

A screenshot of a computer code

Description automatically generated

* Now we predict the classes from the test dataset using our trained model.
* Then we check the accuracy score of the predicted classes.
* Accuracy score() takes true values and predicted values and returns the percentage of accuracy.

**Output:**  
0.9666666666666667

The accuracy is above 96%.

Detailed classification report based on the test dataset

A close-up of a white background

Description automatically generatedA screenshot of a computer

Description automatically generated

A number of numbers on a white background

Description automatically generated

* The classification report gives a detailed report of the prediction.
* Precision defines the ratio of true positives to the sum of true positive and false positives.
* Recall defines the ratio of true positive to the sum of true positive and false negative.
* F1-score is the meaning of precision and recall value.
* Support is the number of actual occurrences of the class in the specified dataset.

Step 5: Testing the Model:

A screenshot of a computer code

Description automatically generated

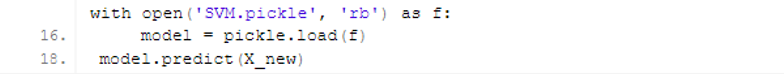
* Here we take some random values based on the average plot to see if the model can predict accurately.

**Output:**

Prediction of Species: [‘Iris-setosa’ ‘Iris-versicolor’ ‘Iris-virginica’]

It looks like the model is predicting correctly because the setosa is shortest and virginica is the longest and versicolor is in between these two.

A white background with black text

Description automatically generated 

* We can save the model using pickle format.
* And again we can load the model in any other program using pickle and use it using model.predict to predict the iris data.

# **Conclusion**

In this project, we learned to train our own supervised machine learning model using Iris Flower Classification Project with Machine Learning. Through this project, we learned about machine learning, data analysis, data visualization, model creation, etc.