**Department of Electrical Engineering**

**Faculty Member:** LE Munadi Sial **Date:** 15-Nov-2023

**Semester:** 7th **Group:**

# CS471 Machine Learning

**Lab 9: Introduction to Sci-kit Learn**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **PLO4 - CLO4** | **PLO4 -CLO4** | **PLO5 -CLO5** | **PLO8 -CLO6** | **PLO9 -CLO7** |
| **Name** | **Reg. No** | **Viva /Quiz / Lab Performance** | **Analysis of data in Lab Report** | **Modern Tool Usage** | **Ethics** | **Individual and Team Work** |
|  |  | **5 Marks** | **5 Marks** | **5 Marks** | **5 Marks** | **5 Marks** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

## Introduction

This laboratory exercise will focus on the Scikit Learn (or SKLearn) library for machine learning implementations in python. Scikit Learn contains many useful functions for fitting models using various machine learning techniques such as linear regression, logistic regression, decision trees, support vector machines, k-means clustering, anomaly detection and more.

## Objectives

The following are the main objectives of this lab:

* Extract and prepare the training and test datasets
* Implement linear regression using Scikit learn
* Implement logistic regression using Scikit learn
* Implement k-means clustering using Scikit learn
* Implement decision trees using Scikit learn

## Lab Conduct

* Respect faculty and peers through speech and actions
* The lab faculty will be available to assist the students. In case some aspect of the lab experiment is not understood, the students are advised to seek help from the faculty.
* In the tasks, there are commented lines such as #YOUR CODE STARTS HERE# where you have to provide the code. You must put the code/screenshot/plot between the #START and #END parts of these commented lines. Do NOT remove the commented lines.
* Use the tab key to provide the indentation in python.
* When you provide the code in the report, keep the font size at 12

**Theory**

Scikit Learn is a python library that contains a wide arsenal of functions pertaining to machine learning. It also contains its own datasets for trying out the machine learning algorithms. Scikit learns API interface can be divided into three types: estimator, predictor and transformer. The estimators are used to fit the model in accordance with some algorithm. The predictors use the fitted model to make prediction on test features. The transformers are used for the conversion of data.

A brief summary of the relevant keywords and functions in python is provided below:

**print()** output text on console

**input()** get input from user on console

**range()**  create a sequence of numbers

**len()** gives the number of characters in a string

**if** contains code that executes depending on a logical condition

**else** connects with **if** and **elif**, executes when conditions are not met

**elif** equivalent to **else if**

**while** loops code as long as a condition is true

**for** loops code through a sequence of items in an iterable object

**break** exit loop immediately

**continue** jump to the next iteration of the loop

**def** used to define a function

**pd.read\_csv** import csv file as a dataframe

**df.to\_csv** export dataframe as a csv file

**Lab Task 1 – Linear Regression \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Download a dataset containing at least 5 feature columns and a label column containing continuous data. Use functions from Sci-kit learn to train a model using linear regression. You will need to split your dataset into training and test portions. Vary the step size and regularization parameters to get at least 6 plots of the training loss and test loss. Lastly, save the weights of the best trained model and use them to make at least five predictions.

Provide the codes and all of the relevant screenshots of your work. Also, give brief explanation of the functions you are using in your codes.

***### TASK 1 CODE STARTS HERE ###***

*### TASK 1 CODE ENDS HERE ###*

***### TASK 1 SCREENSHOTS START HERE ###***

*### TASK 1 SCREENSHOTS END HERE ###*

***### TASK 1 EXPLANATION STARTS HERE ###***

*### TASK 1 EXPLANATION ENDS HERE ###*

**Lab Task 2 – Logistic Regression \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Download a dataset containing at least 5 feature columns and a label column containing discrete data. Use functions from Sci-kit learn to train a model using logistic regression. You will need to split your dataset into training and test portions. Vary the step size and regularization parameters to get at least 6 models of the training. For each model, plot the training loss (vs. epochs), test loss (vs. epochs), precision (vs. epochs) and recall (vs. epochs). Additionally, plot the precision-recall plots for each trained model.

Lastly, save the weights of the best trained model and use them to make at least five predictions. Make a scatter plot for each of your prediction. For this, you will need to show the all of the dataset examples with their labeled classes. Your prediction must be shown as a distinct point in the scatter plots.

Provide the code and all of the relevant screenshots of your work. Also, give brief explanation of the functions you are using in your codes.

***### TASK 2 CODE STARTS HERE ###***

*### TASK 2 CODE ENDS HERE ###*

***### TASK 2 SCREENSHOTS START HERE ###***

*### TASK 2 SCREENSHOTS END HERE ###*

***### TASK 2 EXPLANATION STARTS HERE ###***

*### TASK 2 EXPLANATION ENDS HERE ###*

**Lab Task 3 – K-means Clustering \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Download a dataset containing at least 4 feature columns. Use functions from Sci-kit learn to perform K-means clustering on the following cases:

* 2 features combination
* 3 features combination
* 4 features combination

For each of the above, perform clustering from k = 2 to K clusters (K is up to your choice). For each combination case, make at least 3 cluster plots. Also, make a graph of cost vs. K for all of the 3 combination cases. Use the elbow method to determine the best number of clusters in each case.

Provide the code and all of the relevant screenshots of your work. Also, give brief explanation of the functions you are using in your codes.

***### TASK 3 CODE STARTS HERE ###***

*### TASK 3 CODE ENDS HERE ###*

***### TASK 3 SCREENSHOTS START HERE ###***

*### TASK 3 SCREENSHOTS END HERE ###*

***### TASK 3 EXPLANATION STARTS HERE ###***

*### TASK 3 EXPLANATION ENDS HERE ###*