Lab 1: Introduction to Python libraries for Deep Learning

Objective

Students are expected to gain knowledge and skills related to python language based libraries and frameworks that are needed for deep learning.

Introduction

This lab is prepared to familiarize the libraries that are commonly used in deep learning. In this lab, you will be introduced to the following libraries,

- Numpy

Numpy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with mathematical functions to operate on these arrays.

- Matplotlib

Matplotlib is a library for creating static, animated, and interactive visualizations in Python.

- Pandas

Pandas is a library written for the Python programming language for data manipulation and analysis.

- Seaborn

Seaborn is a Python data visualization library based on matplotlib that provides a high-level interface.

Codes for the lab - https://github.com/jeewaka-p/SE4050-DL-Lecture02

Online store customer dataset - https://www.kaggle.com/datasets/mountboy/online-store-customer-data?resource=download

Au nano particle dataset (google drive) -

https://drive.google.com/drive/folders/1_6-GpmILQB4-8rMf5oW604pP1Nx54dYx?usp=sharing

For this lab, you need to understand the content of the Jupyter notebooks given in the above github repository and then run the codes as well. We suggest you to use google colab to run the notebooks. Each notebook contains a link to a colab implementation. **Make a local copy of the notebook in your google drive and run the codes.** Dataset needed for the pandas_tutorial.ipynb (Online store customer dataset) is given in the above link.

Complete the following tasks and submit the codes as instructed in the submission section below.

Task 1

1. Create a random 4x4 array from an exponential distribution. Refer numpy functions.
2. Create a random 100000x1 array from the same exponential distribution and visualize the distribution histogram along with uniform and normal distributions. Use the code given below and do the necessary changes. Change the number of bins and the plot view to make the visualization better.

```
plt.hist(np.random.rand(100000), density=True, bins=100, histtype="step", color="blue", label="rand")

plt.axis([-2.5, 2.5, 0, 1.1])

plt.legend(loc = "upper left")

plt.title("Random distributions")

plt.xlabel("Value")

plt.ylabel("Density")

plt.show()
```

3. Using matplotlib, plot the Z = X**2 + Y**2 in a 3D plot. Limit the x and y values to [-5, 5].
4. Calculate the pearson standard coefficient and spearman rank correlation for the [HP, Attack, Defense, Sp. Ark, Sp. Def, Speed] features in the seaborn tutorial.ipynb. Visualize the results using heatmaps. Make sure the values are displaced in the heatmap as well.

Task 2

Download the Au_nanoParticle_dataset.csv and load the data of the csv file as a dataframe in a new colab notebook. Do the below subtask with the dataset. Use any of the above libraries if needed.

1. Create a new dataframe by filtering all the columns [i.e., features] except N_total, N_bulk, N_surface and R_avg columns.
2. Display the first 20 samples of this dataframe.
3. Calculate the mean, standard deviation and quartile values for each of the above 4 features.
4. Plot the histogram of each of these features in a 1x4 layout.
5. Visualize the scatter plots and histograms of this dataframe using the pairplot functionality of seaborn library.
6. Add the below code and change it such that,
    a. Plots on the diagonal contains the histogram of each feature along with the kernal density estimation plot.
    b. Plots on the lower half [g.map_lower] contains the bivariate kernal density estimation plot.

```
#new_df is the dataframe containing only the above mentioned 4 features.

g = sns.PairGrid(new_df)
```

g.map_upper(sns.histplot) #bivariate histogram

g.map_diag(sns.histplot)

g.map_lower(sns.kdeplot)

Submission

Upload the completed codes to a github repository. Add the github link to a text file and name the text file by your index number. Upload the text file to the submission link on the courseweb.