**Assignment 1 - Python Basics and Numpy**

Due: Sun Jan 22, 2023 11:59pmDue: Sun Jan 22, 2023 11:59pm

9.5/10 Points

#### **Problem Description**

Create a word-document matrix (A) from a text dataset. If the vocabulary (set of unique words) size is M and the number of documents is N, then the size of this matrix will be M X N. Use numpy data structures to create and manipulate this matrix.

#### **Dataset**

This dataset has 20 short text files. The data is available [here](https://iu.instructure.com/courses/2144074/files/150309242/preview)[Download here](https://iu.instructure.com/courses/2144074/files/150309242/download?download_frd=1).

#### **Output**

1. Show the matrix A: report M and N

2. Visualize the matrix: Use heatmap like visualization to display the word-document matrix (higher frequency should be denoted by a darker shade)

3. Using the matrix compute [TF-IDF (Links to an external site.)](https://en.wikipedia.org/wiki/Tf%E2%80%93idf) scores for each word

4. Using [cosine similarity (Links to an external site.)](https://en.wikipedia.org/wiki/Cosine_similarity) find the 3 most similar documents of "10.txt"

5. Using matrix manipulation and numpy create a new matrix B of size N X N, where Bij will represent the number of common words between document i and j. Note: the diagonal of B should always have the highest value

#### **Submission**

Upload the final code file either in .py or .ipynb format. Make sure you properly comment your code and state all you assumptions.

View Rubric

**Assignment 2 - Neural Network**

Due: Mon Feb 6, 2023 11:59pmDue: Mon Feb 6, 2023 11:59pm

10/10 Points

Download and complete any one of these files. The instructions are in the files. Submit the completed file.

[Assignment2.py](https://iu.instructure.com/courses/2144074/files/150309247/preview)[Download Assignment2.py](https://iu.instructure.com/courses/2144074/files/150309247/download?download_frd=1)

[Assignment2.ipynb](https://iu.instructure.com/courses/2144074/files/150309248/download?wrap=1)

# Lab 1 - Deep Neural Networks using Pytorch/Tensorflow

## Instructions

[Lab 1-Deep Network.ipynb](https://iu.instructure.com/courses/2144074/files/150309299?wrap=1)

**Assignment 3 - Deep Feed Forward Neural Network**

Due: Fri Feb 24, 2023 11:59pmDue: Fri Feb 24, 2023 11:59pm

Instructions:

[Assignment 3.ipynb](https://iu.instructure.com/courses/2144074/files/150309335/download?wrap=1)[Download Assignment 3.ipynb](https://iu.instructure.com/courses/2144074/files/150309335/download?download_frd=1)

(Use this file to complete the assignment and submit it)

Dataset

[train.csv](https://iu.instructure.com/courses/2144074/files/150309382/download)[Download train.csv](https://iu.instructure.com/courses/2144074/files/150309382/download?download_frd=1)

[test.csv](https://iu.instructure.com/courses/2144074/files/150309349/download)

# Lab 2 - Convolutional Neural Network

## Instructions

[Lab 2 - CNN.ipynb](https://iu.instructure.com/courses/2144074/files/150309384?wrap=1)[Download Lab 2 - CNN.ipynb](https://iu.instructure.com/courses/2144074/files/150309384/download?download_frd=1)

Put this image in the same folder as the notebook file

## Assignment 4 - Convolutional Neural Network

Due: Fri Mar 10, 2023 11:59pmDue: Fri Mar 10, 2023 11:59pm

[Assignment4.ipynb](https://iu.instructure.com/courses/2144074/files/150309319/download?wrap=1)[Download Assignment4.ipynb](https://iu.instructure.com/courses/2144074/files/150309319/download?download_frd=1)

Instructions are in the notebook

# Lab 3 - Recurrent Neural Network

## Instructions

[Lab 3 - RNN.ipynb](https://iu.instructure.com/courses/2144074/files/150309341?wrap=1)[Download Lab 3 - RNN.ipynb](https://iu.instructure.com/courses/2144074/files/150309341/download?download_frd=1)

Data files:

[timeseries\_lab3.npy](https://iu.instructure.com/courses/2144074/files/150309323?wrap=1)[Download timeseries\_lab3.npy](https://iu.instructure.com/courses/2144074/files/150309323/download?download_frd=1)

[timeseries\_lab3.csv](https://iu.instructure.com/courses/2144074/files/150309322?wrap=1)[Download timeseries\_lab3.csv](https://iu.instructure.com/courses/2144074/files/150309322/download?download_frd=1)

[elements.txt](https://iu.instructure.com/courses/2144074/files/150309392?wrap=1)[Download elements.txt](https://iu.instructure.com/courses/2144074/files/150309392/download?download_frd=1)

Put this image in the same folder as the notebook file:

[rnn\_langmod.png](https://iu.instructure.com/courses/2144074/files/150309358?wrap=1)

**Assignment 5 - Recurrent Neural Network**

Due: Sun Apr 9, 2023 11:59pmDue: Sun Apr 9, 2023 11:59pm

9/10 Points

#### **Problem Statement #1:**

Build a sequential model to classify names into gender.

Input to the model will be a name, i.e. a sequence of characters.

Use one hot representation of the  characters.

Remove non-ascii characters, if there are any

#### **Outputs:**

Show the effect of the following on the accuracy

1. RNN cells - Simple RNN, LSTM and GRU

2. Dataset size (Randomly select 25%, 50%, 75% and 100% of the data) . For each partial dataset use 80% as training data.

Report overall and class-wise accuracies for all the combinations. (class-wises accuracy should report percentage of correctly predicted male names and female names)

#### **Problem Statement #2:**

Train a language model using these names.

#### **Output**

Generate 100 male names and 100 female names.

Measure the accuracy of classifying these names by using the best-performing model from part 1

#### **Problem Statement #2a:**

Train a language model using names starting with A, M, and Z.

#### **Output**

Generate 50 names

Use perplexity to show the quality of these names, i.e. how realistic these names are

#### **Dataset:**[**name\_gender.csv**](https://iu.instructure.com/courses/2144074/files/150309363/download?wrap=1)

# Lab 4 - Final Lab

[Lab 4 - GAN.ipynb](https://iu.instructure.com/courses/2144074/files/150309324?wrap=1)[Download Lab 4 - GAN.ipynb](https://iu.instructure.com/courses/2144074/files/150309324/download?download_frd=1)

A full example of a GAN implementation is in the notebook