SEIS 763 Machine Learning
Assignment 1
Due: midnight 9/27/21 on Canvas

Individual effort

Create a jupyter notebook called **Assign1.ipynb**. Write code for each of the following questions by having a separate cell for every question. Copy the actual question in a markdown cell and right below that you should have a code cell as shown below.

1) Create a vector of size 15 having third element (i.e. position 2) set to 1 and rest of the elements should be zero.

```
In [ ]: # your code here
```

Import the **numpy** package using the alias **np** and write commands for the following questions. To get credit, you should not be using any loops in your code.

1) Create a vector of size 15 having third element (i.e. position 2) set to 1 and rest of the elements should be zero.

2) Create a vector with values ranging from 21 to 30

3) Reverse the order of elements in the vector in question 2. That is, the previous last element should be the first one now.

4) Define a vector with values ranging from 1 to 12. Now define a 4x3 matrix to store elements from the vector you have defined. That is, first 3 elements of the vector should be row 1 of the matrix, next 3 should be row 2, and so on.

5) Consider the 4x3 matrix from question 4. Add 100 to all values that are greater than equal to 3 and less than equal to 8.

6) Consider the vector x = [1, 0, 6, 9, 10, 0]. Print the indices of non-zero elements.

7) Define a 6x3 matrix with random values. Normalize this matrix by subtracting the mean (of the column) and dividing by the standard deviation (of the column).

8) Define a 3x4 matrix with random values. Define a 4x3 matrix with random values. Multiply the two matrices.

9) Define a 3x3 identity matrix.

10) Define a 3x3 matrix A with random values. Invert the matrix A, let's call this B. Multiply matrices A and B.

For the remaining questions, you will be performing data preparation on the dataset provided. The people.csv dataset contains contact information about people. The attributes of the dataset are *firstname, lastname, age, company name, street name, city, county, state, zipcode, email, url*.

11) Load the dataset into a pandas dataframe and display the first 5 lines of the dataset along with the column headings. Note that the data does not come along with the column headings, so you should be adding that to the data frame.

12) Drop the county column.

13) Keep only those rows that have a minimum of 4 values, otherwise delete them.

14) Delete rows with email missing.

15) Impute the missing values in age with the mean of the column.

16) Perform one hot encoding for the state column.

Submission:

- Make sure each of the cells have been run with the output shown right below. Now, export the notebook as .html file.
- Submit the **.html** file and **.ipynb** notebook on Canvas.