School of Computer Science and Engineering, VIT Chennai.

BCSE209P Machine Learning Lab

Lab-1 Python : Numpy and Pandas

Faculty : Dr. R. Jothi

Due Date : 20/01/2024

Submit your python code (Jupyter notebook): with output for all the questions.

(Use appropriate library functions)

1. Create a 2-D Array (3 rows and 4 cols)

18	21	3	6
5	1	0	11
0	7	0	33

   a. Print indices of 0 (zero) values
   b. Find transpose of the matrix. Also extract 2nd row of the transposed matrix.
   c. Flatten the original matrix to a variable Z and print Z;
   d. Write your original matrix to 'mat.txt' file.

2. Assume marks obtained in three courses by a number of students are available in 'mark.txt' file. Load 'mark.txt' file into a numpy array
   a. print max of mark1, mark2 and mark3
   b. Plot mark1 and mark2 separately using a scatterplot. Also do this for mark1 and mark3. Observe for linearity between the marks and report.
3. Create a dataframe named data with the following entries

	A	B	C
0	1	2	3
1	4	5	6
2	7	8	9

a. Add 5 rows to the dataframe and print;
b. Print rows from 1 to 4 from beginning as well as end (use head and tail functions)
c. Add a column to the dataframe and print
d. Drop the first col and print.

4. Load ‘iris.csv’ file to a dataframe.

a. Print names of all the features of the dataset. Also print shape of the dataset (rows and columns)
b. Find descriptive statistics (min, max, std. deviation) about all the features of the dataset.
c. Count number of distinct elements in species (i.e. the 3 different classes) also count the number of samples in each of three classes
d. Extract all the rows pertaining to 'setosa' into a dataframe named as df_setosa. Similarly extract for 'versicolor', and 'virginica classes.
e. Visualize the correlation between sepal length and sepal width.