

# Introduction to Intelligent Systems, Summer 2022

## Assignment 3

Max Marks: 15

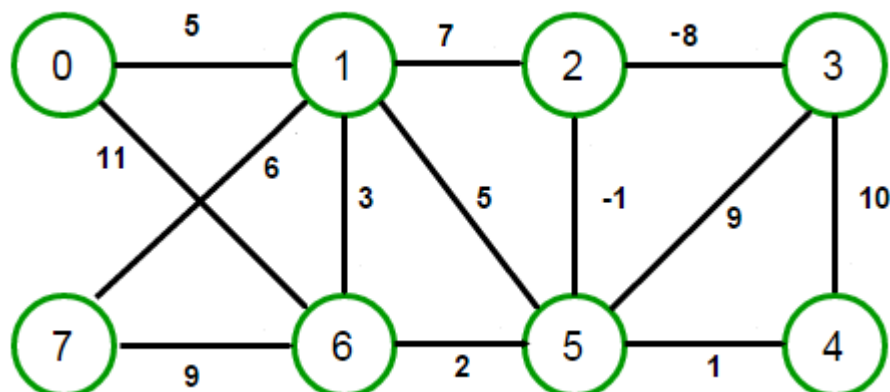
Due Date: 19/07/2022, 11:59 PM

---

### Instructions

- This is an individual assignment.
  - Try to attempt all questions.
  - The questions should be your individual effort. Copying/Plagiarism will be dealt with strictly.
  - No extension days will be permitted.
  - Make a .Zip file (**rollno\_Assignment3.zip**) that contain all the questions. Make sure to number the questions correctly in the .Zip file.
  - Remember to Turn in after uploading on Google Classroom. No excuses or issues would be taken regarding this after the deadline.
- 

**Q1.** In this assignment, you need to find the total cost of traveling from the start node to the end node using breadth-first search and depth-first search using prolog. Given here the start node as node “0” and the end node be “4”. The edge connecting any two nodes depicts the cost between those two nodes: (5)



**Q2.** In the given CSV it is insurance data which consists of the following columns: (5)

- age: age of the primary beneficiary.
- sex: insurance contractor gender, female, male.

- BMI: Body mass index, providing an understanding of the body, weights that are relatively high or low relative to height, an objective index of body weight ( $\text{kg} / \text{m}^2$ ) using the ratio of height to weight, ideally 18.5 to 24.9.
- children: Number of children covered by health insurance / Number of dependents.
- smoker: Smoking.
- region: the beneficiary's residential area in the US, northeast, southeast, southwest, and northwest.
- charges: Individual medical costs billed by health insurance.
- Divide the data into 80:20 ratio and use 80% of the data as training and the rest as testing data.
  - Apply a linear regression model to the following data and predict the accuracy and mean absolute error. You can also apply a few pre-processing to the data to increase the accuracy.

**Q3.** For this question you have to use the Diabetes dataset. (5)  
Divide the data into a 70:30 ratio and use 70% of the data as training and the rest as testing data.

- Visualize and analyze the dataset (use plots to visualize the data).
- Apply a logistic regression model to the following data and predict the accuracy. Print the training and testing accuracy, and confusion matrix for the same.