

Name:	
IIT ID Number:	

CS 579: Online Social Network Analysis

Homework I - Graph Essentials, Data Mining

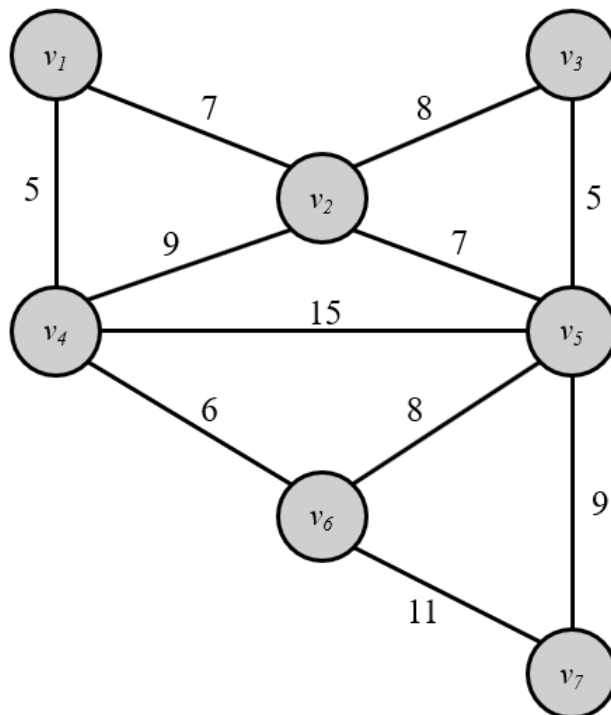
Prof. Kai Shu

Due at 2021 Feb. 7th, 11:59 PM

This is an *individual* homework assignment. Please submit a digital copy of this homework to **Blackboard**. For your solutions, even when not explicitly asked, you are supposed to concisely justify your answers.

1. [Graph Algorithms]

- (a) Compute the shortest path between v_1 and other nodes using Dijkstra's algorithm for the following graph.



Node	Distance from v_1
v_2	
v_3	
v_4	
v_5	
v_6	
v_7	

Name:	
IIT ID Number:	

- (b) In the space below, draw a simple example of a directed graph with negative-weight edges for which Dijkstra's algorithm produces incorrect answers.

- (c) Argue whether “Algorithm 1” below always produces the shortest paths from one source node to others for graphs that have negative weights but do not have negative cycles.

Algorithm 1: Dijkstra Algorithm for graphs with negative weights.

Input : Adjacency Matrix M , Source node s .

Output: Shortest Path from s to other nodes.

1 $C \leftarrow$ Find minimum weight in M

2 **for** all i and j :

3 $M[i, j] \leftarrow M[i, j] - C$

4 **return** Dijkstra(M, s) *// use the original Dijkstra algorithm to find the shortest paths*

2. [Network Algorithms] For a real-world social network, is BFS or DFS more desirable? Why? Provide details.

Name:	
IIT ID Number:	

3. **[Decision Tree and Data Types]** Consider the given dataset below. Answer the following questions:

Instance	Age	Income	Student	Credit Rating	Buy Computer
1	25	High	No	Fair	No
2	20	High	No	Excellent	No
3	32	High	No	Fair	Yes
4	45	Medium	No	Fair	Yes
5	41	Low	Yes	Fair	Yes
6	41	Low	Yes	Excellent	No
7	36	Low	Yes	Excellent	Yes
8	27	Medium	No	Fair	No
9	30	Medium	Yes	Fair	Yes
10	42	Medium	Yes	Fair	Yes
11	29	Medium	Yes	Excellent	Yes
12	31	Medium	No	Excellent	Yes
13	33	High	Yes	Fair	Yes
14	41	Medium	No	Excellent	No

- (a) Specify the data types (Nominal, Ordinal, Interval, Ratio) for each of the four attributes (Age, Income, Student, Credit Rating) in the given data.

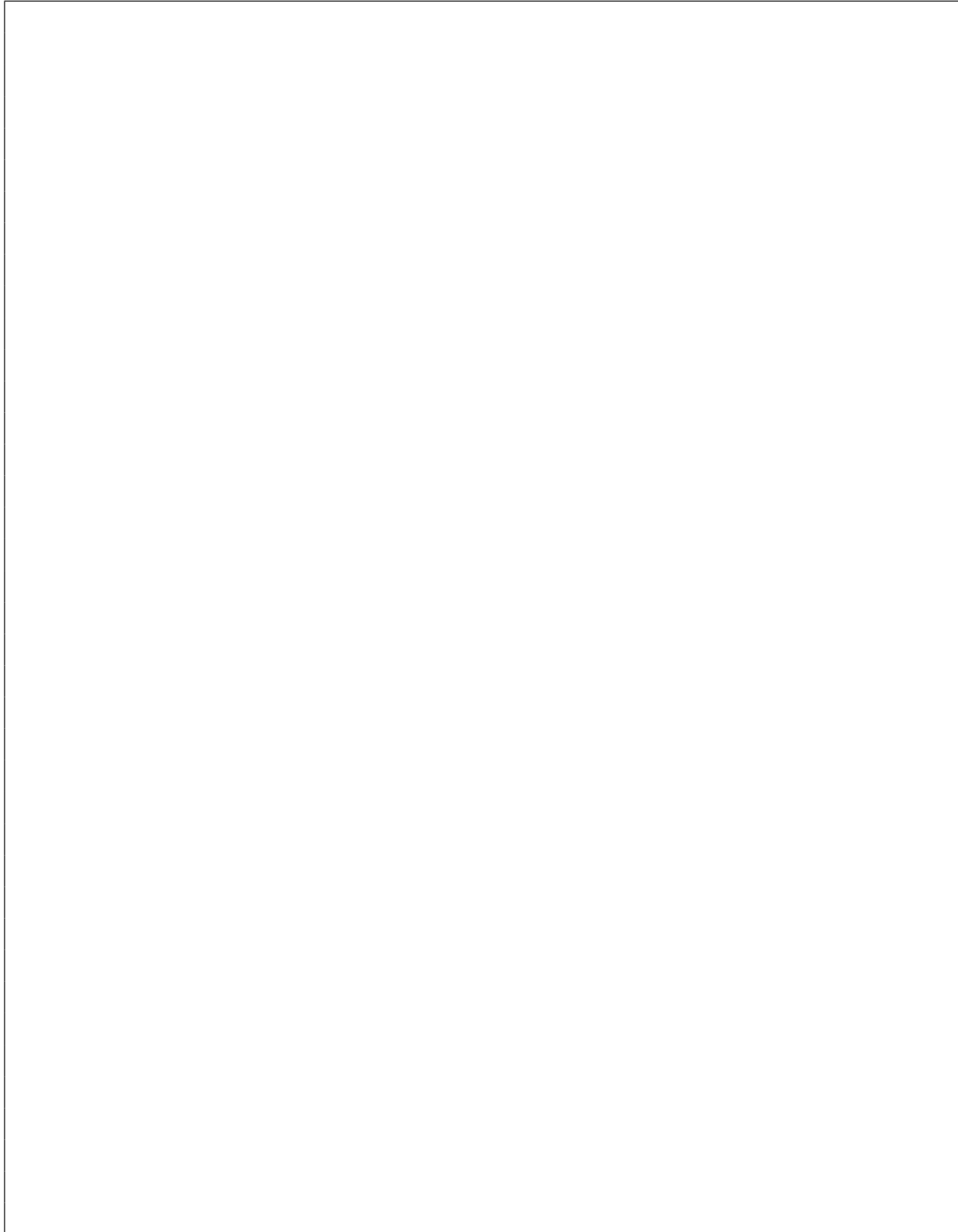
	Age	Income	Student	Credit Rating
Data Type				

- (b) Now assume that we have discretized the real-value “Age” attribute into three categories: 1) 30L: “Age” ≤ 30 , 2) 41H: “Age” ≥ 41 , and 3) BET: $31 \leq \text{“Age”} \leq 40$. What is the new data type for the “Age” attribute given this change?

	Age
Data Type	

Name:	
IIT ID Number:	

- (c) Using the ID3 algorithm that we discussed in the class, generate the decision tree for the given dataset. Assume that “Buy Computer” attribute is the class label and the “Age” attribute is discretized as we discussed in previous question. Note that there could be more than one tree that fits the same data and we only need one! Show all your work for each step in making decision tree and explain how you select decision tree nodes and branches.



Name:	
IIT ID Number:	

4. **[Naive Bayes Classification]** Using the Naive Bayes algorithm and the table given in question 3, what would be the label for the following instance. Assume that “Buy Computer” attribute is the class label and the “Age” attribute is discretized as we discussed in 3.(b).

	Age	Income	Student	Credit Rating	Buy Computer
Instance 15	26	Low	Yes	Fair	?

Name:	
IIT ID Number:	

Notes